THE AMERICAN MATHEMATICAL MONTHLY

(FOUNDED IN 1894 BY BENJAMIN F. FINKEL)

THE OFFICIAL JOURNAL OF

THE MATHEMATICAL ASSOCIATION OF AMERICA, INC.

VOLUME 77 — NUMBER 1

CONTENTS

(*continued on inside cover*)

JANUARY 1970

NOTICE TO AUTHORS

Articles should be typewritten on good quality paper, triple spaced with wide margins. Submit the original and a duplicate; keep a complete copy for protection against loss. Please follow the format used in current issues of the MONTHLY. The manuscript must make the author's intent clear to the printer. A paper accepted for publication by the editor but not acceptable to the printer will be returned to the author for retyping. Please use the American Mathematical Society *Manual for Authors of Mathematical Papers* as your guide and please conform to correct English usage. A *Manual for Monthly Authors* is available on request.

Prospectively MONTHLY authors are advised to consult the Statement of Policy in the January 1969 issue, p. 2. Specialized research is usually unsuitable. Backlog: Main Articles 9 months, Math. Notes 7 months, Research Problems 4 months, Classroom Notes 6 months, Math. Education 4 months.

EDITORIAL CORRESPONDENCE: to HARLEY FLANDERS, American Mathematical Monthly, Purdue University, Lafayette, IN 47907; ADVERTISING CORRESPONDENCE: to RAOUL HAILPERN, Mathematical Association of America, SUNY at Buffalo, Buffalo, N. Y. 14214; CHANGE OF ADDRESS and SUBSCRIPTIONS: to A. B. WILLCOX, Mathematical Association of America, 1225 Connecticut Ave., N.W., Washington, D.C. 20036.

Annual dues for members of the Association (including a subscription to the American Mathematical Monthly) are $10.00. For nonmembers the subscription price is $15.00.

PUBLISHED BY THE ASSOCIATION at Washington, D. C., and Menasha, Wisconsin during the months of January, February, March, April, May, June–July, August–September, October, November, December.

Second-class postage paid at Washington, D. C., and additional mailing offices. Acceptance for mailing at special rate of postage provided for in the Act of February 28, 1925, embodied in Paragraph 4, Section 538, P.L. and R., authorized April 1, 1926.

PRINTED IN THE UNITED STATES OF AMERICA

ERGODIC AUTOMORPHISMS

LOUIS AUSLANDER, City University of New York

Introduction. In recent years a great deal of progress has been made in determining when measure preserving mappings of measure spaces are ergodic. We shall examine some of the new methods developed in [2], [3], and [4] that have enabled us to make this progress. Not only should this serve as an introduction to the subject, but also it should help clarify the power of the results in these papers. This is because many of the computations given here, specializations of the general results, have never been explicitly or even implicitly presented before.

To keep technical problems to a minimum, we have centered our study about automorphisms of tori and an example that is a natural generalization of this concept. The example will show why the classical function theoretic methods cannot yet be generalized. In our final part of this paper, we shall look again at automorphisms of tori, but this time we shall use the most general proof. Thus we shall see how the method of proof has been altered by our need to treat more general problems. To show the success of our method, we shall finish the paper with the study of ergodic properties of affine motions of tori first presented in [5].

Part I. Torus automorphisms.

1. Basic definitions. Let $V^n = \{(x_1, \cdots, x_n)\}$ be the n-dimensional real vector space and let $L = \{(m_1, \cdots, m_n)\}$, m_i an integer. Then L is a subgroup of V^n, called a *lattice subgroup*, and the quotient group $V^n/L = T^n$ is the *n-dimensional torus group*. Let A be a linear transformation of V^n which maps L onto itself. Then relative to the above implicit basis of V^n, A can be represented as a matrix with integer entries and determinant, $|A|$, equal to ± 1. Then A induces an automorphism A_L of T^n onto itself; conversely, if B is any automorphism of T^n, there exists an automorphism A of V^n onto itself which maps L onto itself and such that

$$B = A_L.$$

Let us define a measure on V^n by the nonsingular n-form

$$dx_1 \wedge \cdots \wedge dx_n.$$

Then A induces the mapping dA on n-forms, and we have

$$|(dA)(dx_1 \wedge \cdots \wedge dx_n)| = ||A| dx_1 \wedge \cdots \wedge dx_n| = |dx_1 \wedge \cdots \wedge dx_n|.$$

Therefore A acts as a measure-preserving transformation on the measure space

Professor Auslandei is a University of Chicago student of S. S. Chern. He has taught at a number of universities and been a visitor at the Institute for Advanced Study, Yale, and Berkeley. His books include *Introduction to the Theory of Manifolds*, *Differential Geometry*, *What are Numbers*, and *Calculus*, and he is known for his research on solvable Lie groups. *Editor.*

$(V^n, dx_1 \wedge \cdots \wedge dx_n)$. Since $dx_1 \wedge \cdots \wedge dx_n$ is invariant under translations by L, the n-form $dx_1 \wedge \cdots \wedge dx_n$ projects to an n-form on $V^n/L = T^n$ and determines a measure, μ on T^n. Clearly A_L preserves the measure μ. Thus A_L determines a measure preserving automorphism of the torus T^n.

DEFINITION 1. *Let (X, μ) be a measure space with $\mu(X) < \infty$ and let B be a measure preserving mapping of X onto itself. We say that B acts ergodically on (X, μ) if the only measurable subsets U of X invariant under B, i.e., such that $BU = U$, have the property that*

$$\mu(U) = 0 \quad or \quad \mu(U) = \mu(X).$$

Since if U is measurable, the complement of U is measurable, another way of stating the condition of ergodicity is as follows: B acts ergodically if there are not two disjoint measurable subsets U_1 and U_2 of X invariant under B such that

$$X = U_1 \cup U_2 \quad \text{and} \quad \mu(U_1) > 0 \quad \text{and} \quad \mu(U_2) > 0.$$

2. Nonergodic automorphisms of a torus. It is now a classical result that a necessary and sufficient condition for an automorphism A_L of a torus T^n to be ergodic is that none of the eigenvalues of A be a root of unity. Let us begin by showing geometrically that this condition is necessary.

LEMMA 1. *Let A_L be an automorphism of the torus T^n and let A have an eigenvalue equal to* 1. *Then T^n contains invariant subsets under A_L of arbitrary small positive measure.*

Proof. Let I be the identity operator on V^n and let W be the range of $A - I$. Now W is a proper subset of V^n since $(A-I)$ is singular by hypothesis. Because $(A-I)$ is a matrix with integer entries it maps L into itself. But W is spanned by $(A-I)L$, since L contains a basis of V^n. Thus $W \cap L$ contains a basis of W; hence

$$W/W \cap L$$

is compact.

Now let us choose a basis $\xi_1, \cdots, \xi_n$ of V^n such that $\xi_1, \cdots, \xi_k$ span W. Then (relative to the basis $\xi_1, \cdots, \xi_n$) if we consider A operating in column vectors, we have

$$A = \begin{bmatrix} A_1 & A_2 \\ 0 & I \end{bmatrix},$$

where I is the $(n-k)\times(n-k)$ identity matrix and 0 is the $k\times(n-k)$ trivial matrix. Thus A maps the set U_ϵ defined by

$$U_\epsilon = \{(\xi_1, \cdots, \xi_n) \mid |\xi_i| \leqq \epsilon < C, i = k+1, \cdots, n\}$$

onto itself for all ϵ, when C is sufficiently small. Now

$$U_\epsilon/W \cap L = W/W \cap L \times [-\epsilon, \epsilon] \times \cdots \times [-\epsilon, \epsilon]$$

(where there are $n-k$ factors $[-\epsilon, \epsilon]$) is compact. The image U_ϵ^* of U_ϵ in T^n is

compact and has measure less than $K\epsilon^{n-k}$, where K is a fixed constant. This proves our assertion.

THEOREM 2. *Let A_L be an automorphism of the torus T^n such that A has an eigenvalue which is a root of unity. Then A_L is not ergodic on T^n.*

Proof. Let α be an integer such that A^α has an eigenvalue 1, and let U_ϵ^* be an invariant space for A^α as in Lemma 1. Then $U_{i=0}^{\alpha-1}A^i(U_\epsilon^*)$ is an invariant set for A of positive measure D, where

$$K\epsilon^{n-k} \leqq D \leqq \alpha K\epsilon^{n-k}.$$

By choosing ϵ sufficiently small we can prove our assertion.

3. Ergodic automorphisms of the torus. Before proving the converse of Theorem 2 it will be convenient to look at the concept of ergodicity from another point of view.

Let (X, μ) be a measure space, $\mu(X) < \infty$, and let B be a measure preserving transformation of X. Let $L^2(X, \mu)$ be the Hilbert space of square summable functions on X, and let

$$U_B(f)(x) = f(B(x)) \qquad f \in L^2(X, \mu), \qquad x \in X.$$

Then it is easy to verify that U_B is a unitary operator on $L^2(X, \mu)$.

THEOREM 3. *A necessary and sufficient condition for a measure-preserving transformation B to be ergodic on (X, μ) is that*

$$U_Bf = f$$

implies $f =$ constant a.e.

Proof. Let B not be ergodic on X. Then there exist two disjoint measurable sets U_1 and U_2 each of positive measure and each invariant under B such that

$$X = U_1 \cup U_2.$$

If χ is the characteristic function of U_1, then $U_B(\chi) = \chi$ since $B(U_1) = U_1$. But $\chi \neq$constant a.e.

Conversely, let $U_Bf = f$, $f \in L^2(X, \mu)$, $f \neq$constant a.e. Then there exists a set of measure zero, say K, such that

$$f(x) = f(Bx) \qquad x \in X - K.$$

Let us choose C_0 such that

$$f^{-1}\{r \mid r < C_0\} = U_1 \quad \text{and} \quad f^{-1}\{r \mid C_0 \leqq r\} = U_2$$

have positive measure. Then both $U_1 - (U_1 \cap K)$ and $U_2 - (U_2 \cap K)$ have positive measure and are invariant under B. Thus B is not ergodic. These results prove our assertion.

THEOREM 4. *Let A_L be an automorphism of the torus T^n with A having no eigenvalues which are roots of unity. Then A_L is ergodic.*

Proof. Let $\ell \in \boldsymbol{L}$, let $x \in \boldsymbol{V}^n$, and let $\ell \cdot x$ denote the dot product in $\boldsymbol{V}^n$. Then the functions $e^{2\pi i \ell \cdot x}$ can be viewed first as functions on $\boldsymbol{T}^n$ and second as an orthonormal basis for $L^2(\boldsymbol{T}^n)$. Thus if $f \in L^2(\boldsymbol{T}^n)$, then

$$f = \sum_{\ell \in L} b_\ell e^{2\pi i \ell \cdot x}.$$

We next observe that

$$U_{A_L}(e^{2\pi i \ell \cdot x}) = e^{2\pi i \ell \cdot (Ax)} = e^{2\pi i (A'\ell) \cdot x},$$

where A' is the transpose of A acting on the lattice $\boldsymbol{L}$. Hence

$$U_{A_L} f = \sum_{\ell \in L} b_\ell e^{2\pi i \ell \cdot A(x)} = \sum_{\ell \in L} b_\ell e^{2\pi i (A'\ell) \cdot x}.$$

Since A' is a 1-1 mapping of L onto itself, setting $B = (A')^{-1}$ we have

$$U_{A_L} f = \sum_{\ell \in L} b_{B(\ell)} e^{2\pi i \ell \cdot x}.$$

Thus $U_A f = f$ is equivalent to $b_{B(\ell)} = b_\ell$.

Now B acting on L cannot have a finite orbit provided $\ell \neq 0$, i.e., the points

$$\ell, B\ell, \cdots, B^m\ell, \cdots$$

are all distinct. For if $B^r\ell = B^s\ell$, where $r > s$, then

$$(B^{r-s} - 1)(B^s(\ell)) = 0, \tag{1}$$

and B^{r-s} has an eigenvalue equal to 1. But B has an eigenvalue which is a root of unity if and only if A has an eigenvalue which is a root of unity. Thus (1) would contradict the fact that A has no eigenvalues that are roots of unity.

Suppose A is not ergodic. By Theorem 3, there is a function $f \in L^2(\boldsymbol{T}^n)$ such that $U_A f = f$ and $f \neq$ constant a.e. Thus there exists an $\ell_0 \neq 0$ such that $b_{\ell_0} \neq 0$. But then

$$\sum_{\ell \in L} b_\ell^2 \leqq \int_{T^n} |f(x)|^2 dx < \infty$$

and

$$\sum_{\ell \in L} b_\ell^2 \geqq b_{\ell_0}^2 + b_{B\ell_0}^2 + \cdots + b_{B^m\ell_0}^2 + \cdots = \infty$$

which is impossible. Thus Theorem 4 has been proved.

Thus Theorems 2 and 4 together have established the following: a necessary and sufficient condition for an automorphsim A_L of a torus to be ergodic is that A has no eigenvalues which are roots of unity.

Part II. Generalizations of tori

4. Basic definitions. There exists a natural generalization of the torus which involves non-abelian nilpotent Lie groups. Rather than looking at this in its

full generality, we shall content ourselves with a detailed examination of the simplest special case.

Let x_1, x_2, x_3 be real numbers and consider the set of matrices

$$N_3 = \left\{ \begin{bmatrix} 1 & x_1 & x_3 \\ 0 & 1 & x_2 \\ 0 & 0 & 1 \end{bmatrix} \right\}.$$

Then N_3 is a group, since

$$\begin{bmatrix} 1 & x_1 & x_3 \\ 0 & 1 & x_2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & y_1 & y_3 \\ 0 & 1 & y_2 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & x_1 + y_1 & x_3 + y_3 + x_1 y_2 \\ 0 & 1 & x_2 + y_2 \\ 0 & 0 & 1 \end{bmatrix},$$

and

$$\begin{bmatrix} 1 & x_1 & x_3 \\ 0 & 1 & x_2 \\ 0 & 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & -x_1 & -x_3 + x_1 x_2 \\ 0 & 1 & -x_2 \\ 0 & 0 & 1 \end{bmatrix}.$$

If we denote the element

$$\begin{bmatrix} 1 & x_1 & x_3 \\ 0 & 1 & x_2 \\ 0 & 0 & 1 \end{bmatrix}$$

of N_3 by the triple (x_1, x_2, x_3) then the group multiplication becomes

$$(x_1, x_2, x_3)(y_1, y_2, y_3) = (x_1 + y_1, x_2 + y_2, x_3 + y_3 + x_1 y_2);$$

thus instead of having a linear rule for multiplying group elements we have a polynomial rule.

We define a measure on N_3 by means of the nonsingular form $dx_1 \wedge dx_2 \wedge dx_3$. We shall see that this form on N_3 is invariant under N_3 acting on itself by left multiplication. Since

$$(a_1, a_2, a_3)(x_1, x_2, x_3) = (x_1 + a_1, x_2 + a_2, x_3 + a_3 + a_1 x_2),$$

we have

$$\begin{aligned} dx_1 \wedge dx_2 \wedge dx_2 &\to d(x_1 + a_1) \wedge d(x_2 + a_2) \wedge d(x_3 + a_3 + a_1 x_2) \\ &= dx_1 \wedge dx_2 \wedge (dx_3 + a_1 dx_2) = dx_1 \wedge dx_2 \wedge dx_3. \end{aligned}$$

A similar computation shows that the form $dx_1 \wedge dx_2 \wedge dx_3$ is right invariant under N_3 acting on itself. Let us denote this measure space by (N_3, dx). Let

$$L = \left\{ (m_1, m_2, \tfrac{1}{2} m_3) \in N_3 \mid m_i \text{ integers } i = 1, 2, 3 \right\}.$$

Then it is straightforward to verify that L is a subgroup of N_3, so we may form

the homogeneous space N_3/L of right cosets nL, $n \in N_3$. Since L is *not* normal in N_3, the space N_3/L is not a group. But it has several important properties that follow from the lemma below.

LEMMA 5. *Let N_3 and L be as above, and let*

$$C = \{n = (x_1, x_2, x_3) \in N_3 \mid 0 \leqq x_1, x_2 < 1 \text{ and } 0 \leqq x_3 < \tfrac{1}{2}\}.$$

Then: (1) *For $n_1, n_2 \in C$ there exists no $\ell \in L$ such that $n_1\ell = n_2$.* (2) *For any $n \in N_3$ there exists $n_1 \in C$ and $\ell \in L$ such that $n\ell = n_1$.*

Proof. Let (x_1, x_2, x_3) and $(y_1, y_2, y_3) \in C$, and let $(m_1, m_2, \frac{1}{2}m_3) \in L$ satisfy

$$(x_1, x_2, x_3)(m_1, m_2, \tfrac{1}{2}m_3) = (y_1, y_2, y_3).$$

Since $(x_1, x_2, x_3)(m_1, m_2, \frac{1}{2}m_3) = (x_1+m_1, x_2+m_2, x_3+\frac{1}{2}m_3+x_1m_2)$, the restrictions $0 \leqq y_i < 1$ and $0 \leqq x_1, x_2 < 1$ imply that $m_1 = m_2 = 0$. Also $0 \leqq x_3+\frac{1}{2}m_3 < \frac{1}{2}$ with $0 \leqq x_3 < \frac{1}{2}$ and m_3 an integer. This implies that $m_3 = 0$ and proves (1).

To prove assertion (2) let $n = (y_1, y_2, y_3)$. Let $y_i = x_i + m_i$, where m_i is an integer and $0 \leqq x_1, x_2 < 1$. Then

$$(x_1, x_2, x_3)(-m_1, -m_2, 0) = (x_1 - m_1, x_2 - m_2, x_3 - x_1m_2).$$

Let $x_3 - x_1m_2 = \frac{1}{2}m_3 + y_3$ for $0 \leqq x_3 < \frac{1}{2}$ and m_3 an integer. Then

$$(x_1, x_2, x_3)(-m_1, -m_2, -\tfrac{1}{2}m_3) = (y_1, y_2, y_3) \quad \text{and} \quad (x_1, x_2, x_3) \in C.$$

The set C is often called a *fundamental domain* for the group L acting on the space N_3. The existence of C has certain important implications which we list below:

P.1. N_3/L is a differentiable manifold with the natural covering mapping

$$P: N_3 \to N_3/L$$

a differentiable mapping and a local diffeomorphism.

P.2. N_3/L is compact. (This follows from the fact that the closure $\bar{C}$ of C in N_3 is compact and P maps $\bar{C}$ onto N_3/L by Lemma 5.)

P.3. Let A be an automorphism of N_3 which maps L onto itself. Then A determines a 1-1 mapping of N_3/L onto itself which we call an *automorphism* of N_3/L and denote by A_L.

The left and right invariance of the form $dx_1 \wedge dx_2 \wedge dx_3$ on N_3 has also several important consequences:

P.4. P induces a mapping of the 3-form $dx_1 \wedge dx_2 \wedge dx_3$ to a 3-form on N_3/L and thus determines a measure on N_3/L which we denote by μ.

P.5. N_3 acts on N_3/L by $n_0(nL) = (n_0n)L$, $n_0 \in N_3$.

This defines N_3 acting on N_3/L, and it acts as a group of measure preserving transformations on the measure space $(N_3/L, \mu)$.

5. Automorphisms of N_3 and N_3/L. In order to facilitate the discussion of automorphisms of N_3 and N_3/L, it is very convenient to introduce $L(N_3)$, the

Lie algebra of N_3. Since we shall not need the algebra structure of $L(N_3)$, we shall content ourselves with defining the vector space structure on $L(N_3)$.

Consider the space $E(V^3)$ of all 3×3 real matrices as a 9-dimensional real euclidean space. We can consider N_3 as a 3-dimensional submanifold of $E(V^3)$. Hence, we may talk about the tangent space of N_3 at any of its points, n, and denote this by $T_n(N_3)$, $n\in N_3$. Now the identity e is an element of N_3, and so we may consider $T_e(N_3)$. If we consider $T_e(N_3)$ as a subspace of $E(V^3)$, it is precisely the subspace

$$\begin{bmatrix} 0 & a & c \\ 0 & 0 & b \\ 0 & 0 & 0 \end{bmatrix},$$

where a, b, c are real numbers. Now every differentiable mapping of N_3 onto itself which leaves e fixed induces a linear transformation of $T_e(N_3)$ which is the Jacobian of the mapping at e. Thus if A is an automorphism of N_3, then A determines a linear transformation of $T_e(N_3)$ onto itself, which we denote by $ad(A)$.

We next observe that there is a diffeomorphism, exp, of $T_e(N_3)$ onto N_3 defined by

$$\exp\begin{bmatrix} 0 & a & c \\ 0 & 0 & b \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & a & c \\ 0 & 0 & b \\ 0 & 0 & 0 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 0 & a & c \\ 0 & 0 & b \\ 0 & 0 & 0 \end{bmatrix}^2 + $$

$$= \begin{bmatrix} 1 & a & c+\frac{1}{2}ab \\ 0 & 1 & b \\ 0 & 0 & 1 \end{bmatrix}$$

since

$$\begin{bmatrix} 0 & a & c \\ 0 & 0 & b \\ 0 & 0 & 0 \end{bmatrix}^n = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad (n \geqq 3).$$

Thus $\exp(a, b, c) = (a, b, c+\frac{1}{2}ab)$. It is straightforward to verify that exp has an inverse $\log: N_3\to T_e(N_3)$ given by

$$\log\begin{bmatrix} 1 & x_1 & x_3 \\ 0 & 1 & x_2 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & x_1 & x_3-\frac{1}{2}x_1x_2 \\ 0 & 0 & x_2 \\ 0 & 0 & 0 \end{bmatrix}.$$

It is interesting to note that $\log L$ is a lattice in the vector space $T_e(N_3)$. Let us examine one other important property of the exponential mapping.

LEMMA 6. *Consider the* 1*-dimensional vector subspace* V^1

$$(at, bt, ct) \qquad -\infty < t < \infty$$

of $T_e(N_3)$. *Then* $\exp(V^1)$ *is a subgroup of* N_3 *isomorphic to the real line.*

Proof. We have $\exp(at, bt, ct) = (at, bt, ct + \frac{1}{2}abt^2)$. We next note that

$$\begin{aligned}(at_1, bt_1, ct_1 + \tfrac{1}{2}abt_1^2)&(at_2, bt_2, ct_2 + \tfrac{1}{2}abt_2^2) \\ &= (a(t_1 + t_2), b(t_1 + t_2), c(t_1 + t_2) + \tfrac{1}{2}ab(t_1^2 + t_2^2) + abt_1t_2) \\ &= (a(t_1 + t_2), b(t_1 + t_2), c(t_1 + t_2) + \tfrac{1}{2}ab(t_1 + t_2)^2).\end{aligned}$$

Thus $\exp(V^1)$ is a subgroup of N_3 isomorphic to the reals.

Nontrivial subgroups of Lie groups that are continuous images of the line are called *one-parameter groups*. Thus $\exp(V^1)$ is a 1-parameter subgroup of N_3. The converse of Lemma 6 is also true, i.e., every 1-parameter subgroup of N_3 is $\exp(V^1)$ for some 1-dimensional subspace of $T_e(N_3)$. We neither need nor prove this statement.

Let A^* be a linear transformation of $T_e(N_3)$ such that (using column vectors)

$$A^* = \begin{bmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ 0 & 0 & \alpha \end{bmatrix},$$

where $a_{11}, a_{12}, a_{21}, a_{22}$ are integers and

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = \alpha = \pm 1.$$

Then it is a straightforward, if somewhat lengthy, computation to verify the following: If we define $A : N_3 \to N_3$ by

$$A = \exp \circ A^* \circ \log,$$

then A is an automorphism of N_3 which maps L onto itself. Also, each automorphism A of N_3 which maps L onto itself has the property that $ad(A) = A^*$.

Thus the mappings exp and log enable us to pass back and forth from the group N_3 to the vector space $T_e(N_3)$ and replace automorphisms of N_3 by certain linear transformations of $T_e(N_3)$. Now

$$\exp^*(da \wedge db \wedge dc) = da \wedge db \wedge d(c + \tfrac{1}{2}ab) = da \wedge db \wedge dc.$$

Since A^* preserves the form $da \wedge db \wedge dc$ it is straightforward to verify that each automorphism of N_3 of the form

$$A = \exp \circ A^* \circ \log$$

preserves the form $dx_1 \wedge dx_2 \wedge dx_3$ and hence the measure μ. Thus every automorphism of N_3 which maps L onto itself is measure preserving.

DEFINITION. *Let A be an automorphism of N_3 which maps the group L onto itself. Then A induces a measure-preserving diffeomorphism A_L of N_3/L onto itself which we call an automorphism of N_3/L.*

We shall show the following: If the eigenvalues of A are 1, e^{λ}, $e^{-\lambda}$, with real λ, then A_L induces an ergodic automorphism of N_3/L. (The action of A_L is ergodic even if λ is not real, but the proof of this assertion requires a technique that would obscure the main point of our discussion.) The proof of the assertion with real λ will be our goal in the next three sections. Before going on to the next sections, let us prove one important group theoretic lemma.

LEMMA 7. *Let all notation be as above and let v_1 and v_2 be eigenvectors of A^* in $T_e(N_3)$ such that*

$$A^*v_1 = e^{\lambda}v_1, \qquad A^*v_2 = e^{-\lambda}v_2.$$

Then N_3 is generated as a group by the 1-parameter subgroups $\exp(tv_i) = q_i(t)$, $i = 1, 2$. *Furthermore* $A(q_1(t)) = q_1(e^{\lambda}t)$ and $A(q_2(t)) = q_2(e^{-\lambda}t)$.

Proof. The last statements in the lemma follow trivially from our definition of $q_i(t)$ $i = 1, 2$ and the fact that $A = \exp \circ A^* \circ \log$. It remains only to prove the first part. The first part follows from the following general group theoretic facts: (1) The subgroup $Z = \{(0, 0, x_3) \in N_3 \mid \text{all real } x_3\}$ is the commutator subgroup and the center of N_3. (2) The images of $q_1(t)$ and $q_2(t)$ in N_3/Z generate this two-dimensional vector space as a group since they are linearly independent. (3) Any element in Z can be written as the commutator of some pair of elements a, b, where $a \in q_1(t)$ and $b \in q_2(t)$.

6. Flows and automorphisms. Since we do not have a basis of $L^2(N_3/L)$ on which A acts in a controllable fashion, we must find a different way of proving our theorem. It turns out that this is most conveniently done by changing from a discrete to a continuous problem. We shall first examine this in an abstract setting, and then see that in our particular case it can be given a nice formulation.

Let (X, μ) be a finite measure space, and let B be a measure preserving 1-1 mapping of X. We shall build a new measure space $(Y, \mu \times dx)$ on which the reals, R, acts as a group of measure-preserving transformations, and where R acts ergodically on $(Y, \mu \times dx)$ if and only if B acts ergodically on (X, μ). This construction is often called the *skew product* construction.

DEFINITION. *Let $Y_1 = X \times I$, where $I = [0, 1]$, and let dx be the Lebesque measure on I. Take $\mu \times dx$ as the measure on Y_1. Then $X \times 0$ and $X \times 1$ have measure zero in $X \times I$. Now define Y as Y_1 with $X \times 0$ identified with $X \times 1$ by the mapping B. To be precise*

$$Y = \{(x, a) \mid x \in X, \quad 0 \leqq a \leqq 1, \quad \text{and} \quad (x, 0) = (Bx, 1)\}.$$

Then $\mu \times dx$ defines a measure on Y.

Now define the action of the group R on Y. Let R be parameterized by $-\infty < t < \infty$; define $t(y)$ for $y \in Y$ as follows:

If $0 < m \leqq t < m+1$, let

$$t(x, a) = (B^m x, (a + t) \bmod 1) \qquad \text{if } a + t \bmod 1 < 1$$

$$t(x, a) = (B^{m+1} x, (a + t) \bmod 1) \qquad \text{if } a + t \bmod 1 \geqq 1,$$

where m is an integer. We call R acting on the measure space $(Y, \mu \times dx)$ the *skew product flow*.

It is easy to see that R acts as a group of measure preserving transformations on $(Y, \mu \times dx)$. Just as we can talk of a single measure-preserving transformation acting ergodically on a measure space so can we talk of a group of measure-preserving transformations acting ergodically on a measure space.

DEFINITION. *Let (X, μ) be a measure space and let $\mu(X) < \infty$. Let G be a group of measure preserving transformations of X. We say that G acts ergodically on X if the only measurable subsets U of X such that $g(U) = U$ all $g \in G$ have the property that*

$$\mu(U) = 0 \quad \textit{or} \quad \mu(U) = \mu(X).$$

Of course, this definition can be reformulated as follows: G acts ergodically if there are not two disjoint measurable subsets U_1 and U_2 of X such that $g(U_1) = U_1$ and $g(U_2) = U_2$ for all $g \in G$, and with the additional properties

$$X = U_1 \cup U_2, \qquad \mu(U_1) > 0, \qquad \mu(U_2) > 0.$$

The following theorem is the reason for introducing skew product flows.

THEOREM 8. *Let (X, μ) be a measure space with $\mu(X) < \infty$ and let B be a measure preserving mapping of X. The mapping B acts ergodically on (X, μ) if and only if the reals R act ergodically as the skew product flow on $(Y, \mu \times dx)$.*

Proof. Assume B does not operate ergodically on X, i.e., let U be an invariant subset of X such that $0 < \mu(U) < \mu(X)$. Then $U \times [0, 1]$ is invariant in Y and clearly $0 < (\mu \times dx)(U \times [0, 1]) < (\mu \times dx)(Y)$. Thus R does not act ergodically on Y.

Conversely assume R does not act ergodically on Y. Let $V \subset Y$ be invariant under R and satisfying

$$0 < (\mu \times dx)(V) < (\mu \times dx)(Y).$$

Then the pre-image of V in Y_1 is a set of the form $U \times [0, 1]$, where U is invariant under B. But clearly

$$0 < \mu(U) < \mu(X),$$

so B is not ergodic.

Let us now return to our special case of an automorphism A_L acting on N_3/L, and let A be the automorphism of N_3 which determines A_L. Further let

$A = \exp \circ A^* \circ \log$, where A^* is a linear transformation of $T_e(N_3)$. Recall that we are assuming that

$$A^* = \begin{bmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

and that the matrix

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

has positive real eigenvalues. This implies that $A(0, 0, x_3) = (0, 0, x_3)$, since $\exp(0, 0, Ct) = (0, 0, Ct)$, $C \neq 0$ and

$$A^*(0, 0, Ct) = (0, 0, Ct).$$

Let a basis v_1, v_2, v_3 of $T_e(N_3)$ be chosen so that

$$A^*(v_1) = e^{\lambda} v_1, \qquad A^*(v_2) = e^{-\lambda} v_2, \qquad A^*(v_3) = v_3,$$

where $v_3 = (0, 0, C)$ where $C \neq 0$.

We may define a 1-parameter group $A^*(t)$ acting on $T_e(N_3)$ by

$$A^*(t)(v_1) = e^{\lambda t} v_1, \qquad A^*(t)(v_2) = e^{-\lambda t} v_2, \qquad A^*(t)(v_3) = v_3.$$

Let $A(t) = \exp \circ A^*(t) \circ \log$. Then we may verify that $A(t)$ is a 1-parameter group of automorphisms of N_3.

We shall now introduce an algebraic construction that we shall ultimately see is equivalent to the skew product construction.

We begin by defining the semidirect product of two groups A and B, where each element of A may be considered an automorphism of B. The *semidirect product* of A and B, denoted by $A \cdot B$, is the set $A \times B$ with the multiplication given by

$$(a_1, b_1)(a_2, b_2) = (a_1 a_2, \overset{-1}{a_2}(b_1) \cdot b_2).$$

Under this law of composition one can verify that $A \cdot B$ is a group. Note that A and B can be viewed as subgroups of $A \cdot B$ in an obvious way.

Using this construction, we form the groups $A(t) \cdot N_3$ and $A(m) \cdot L$, where m takes on integer values. It is clear that

$$A(t) \cdot N_3 \supset A(m) \cdot L.$$

We introduce the notation $\Gamma = A(m) \cdot L$. Let us now examine the relation of $A(t) \cdot N_3$ to the subgroup Γ.

LEMMA 9. *Let all notation be as defined above. Then Γ is a discrete subgroup of $A(t) \cdot N_3$, and*

$$A(t) \cdot N_3 / \Gamma$$

is compact and homeomorphic to the skew product space determined by $\boldsymbol{N}_3/\boldsymbol{L}$ *and the mapping* A.

Proof. Since $A(t)\cdot\boldsymbol{N}_3$ consists of all pairs (t, n), $t\in\boldsymbol{R}$, $n\in\boldsymbol{N}_3$ we have that (t, x_1, x_2, x_3) (real parameters) serves as a coordinate system for $A(t)\cdot\boldsymbol{N}_3$. Further $\boldsymbol{\Gamma}$ is the subgroup consisting all the points $(m_1, m_2, m_3, \frac{1}{2}m_4)$, where all m_i are integers. Thus $\boldsymbol{\Gamma}$ is a discrete subgroup of $A(t)\cdot\boldsymbol{N}_3$. We may argue essentially as in Lemma 5 to see that the set

$$C' = \left\{(t, x_1, x_2, \tfrac{1}{2}x_3) \in A(t)\cdot\boldsymbol{N}_3 \,\middle|\, 0 \leqq t < 1,\, 0 \leqq x_i < 1,\ i = 1, 2, 3\right\}$$

is a fundamental domain for $\boldsymbol{\Gamma}$ acting on $A(t)\cdot\boldsymbol{N}_3$. Thus again we have properties P1 and P2 holding for $A(t)\cdot\boldsymbol{N}_3/\boldsymbol{\Gamma}$. In particular $A(t)\cdot\boldsymbol{N}_3/\boldsymbol{\Gamma}$ is compact.

It remains to verify our last assertion. This is best carried out in the language of fiber bundles. The reader who is not familiar with this language can convince himself of our assertion by checking the identifications that $\boldsymbol{\Gamma}$ makes on the boundary of C'. In particular $\boldsymbol{L}$ maps the sets

$$\left\{(t_0, x_1, x_2, x_3)\right\}$$

onto themselves, and the identification space is $\boldsymbol{N}_3/\boldsymbol{L}$. Further $A(1)$ maps the set $\{(0, x_1, x_2, x_3)\}$ onto $\{(1, x_1, x_2, x_3)\}$ exactly as A does. Thus on the two boundaries where $t=0$ and $t=1$ of C', the identification is the same as A_L. This shows that $A(t)\cdot\boldsymbol{N}_3/\boldsymbol{L}$ is the desired skew product space.

A more rigorous proof can be based on the following remarks:

$$\begin{array}{c} A(t)\cdot\boldsymbol{N}_3/\boldsymbol{\Gamma} \\ \downarrow \\ A(t)\cdot\boldsymbol{N}_3/\boldsymbol{\Gamma}\boldsymbol{N}_3 \end{array}$$

is a fiber bundle with fiber $\boldsymbol{N}_3/\boldsymbol{L}$ and base space $A(t)/A(m) = A(t)\cdot\boldsymbol{N}_3/\boldsymbol{\Gamma}\boldsymbol{N}_3$, which is topologically a circle. Thus $A(t)\cdot\boldsymbol{N}_3/\boldsymbol{\Gamma}$ may be pictured as

$$\boldsymbol{N}_3/\boldsymbol{\Gamma} \times [0, 1]$$

with $\boldsymbol{N}_3/\boldsymbol{\Gamma}\times 0$ and $\boldsymbol{N}_3/\boldsymbol{\Gamma}\times 1$ identified by the mapping $A(1)$ on the fiber. This again shows that $A(t)\cdot\boldsymbol{N}_3/\boldsymbol{L}$ is the desired skew product space.

LEMMA 10. *Let all notation be as before. Then* $A(t)$ *acts on* $(A(t)\cdot\boldsymbol{N}_3/\boldsymbol{\Gamma}, \mu\times dx)$ *as the skew product flow, where*

$$A(t)(g\boldsymbol{\Gamma}) = (A(t)g)\boldsymbol{\Gamma} \qquad g \in A(t)\cdot\boldsymbol{N}_3.$$

Proof. That $A(t)$ acts on $(A(t)\cdot\boldsymbol{N}_3/\boldsymbol{\Gamma}, \mu\times dx)$ as a group of measure preserving transformations we accept without proof. We shall, however, verify that the action is correct.

Let $(t, n)\in C'$. Once we verify that $A(t)$ acts correctly on these points we shall have verified it in general. Now let $(s, 0)\in A(t)$. Then

$$(s, 0)(t, n) = (s + t, n)$$

by our definition of group multiplication for the semidirect product. Thus for $0 \leqq s+t<1$, we have $(s, 0)(t, n) \in C'$ and the action of $A(s)$ agrees with s in the skew product flow. For $s+t=1$ we have

$$(s+t, n)(-1, 0) = (0, A(1)n) = (0, A(n)).$$

But this also agrees with the skew product flow. The most general case then follows since both actions are group actions. Theorem 7 and Lemmas 8 and 9 combine to yield a proof of the following theorem.

THEOREM 11. *The transformation A acts ergodically on N_3/Γ if and only if $A(t)$ acts ergodically on $(A(t)\cdot \boldsymbol{N}_3/\boldsymbol{\Gamma}, \mu\times dx)$.*

7. Mautner phenomena. There are certain Lie groups whose unitary representations have a strange property which we shall call the Mautner phenomenon. Since this will play a crucial role in our discussion, we shall go into some detail on the matter.

Let $\boldsymbol{H}$ be a separable Hilbert space, and let $\boldsymbol{U}(\boldsymbol{H})$ denote the group of unitary operators on $\boldsymbol{H}$ with the strong operator topology. Let G be a Lie group. Then a continuous homomorphism $U: G \to U(\boldsymbol{H})$ given by $U(g) = U_g$ is called a *unitary representation* of G.

DEFINITION. *Let $(G, p(t))$ be a pair consisting of a Lie group G and a 1-parameter subgroup $p(t)$. Let U be any unitary representation of G and let $\psi \in \boldsymbol{H}$ satisfy $U_{p(t)}\psi = \psi$ for all t. If this implies that $U_g\psi = \psi$ for all $g \in G$, we shall say that $(G, p(t))$ exhibits the Mautner phenomenon.*

The remarkable fact is that there are many Lie groups G which have 1-parameter subgroups $p(t)$ such that the pair $(G, p(t))$ exhibits the Mautner phenomenon. We need to discuss only two such examples.

Example 1. Let S_1 be the group consisting of ordered pairs of real numbers (x_1, x_2) with multiplication given by

$$(x_1, x_2)(y_1, y_2) = (x_1 + y_1, e^{y_1}x_2 + y_2).$$

This is isomorphic to the group of matrices of the form

$$\begin{bmatrix} e^{x_1} & x_2 \\ 0 & 1 \end{bmatrix} \qquad -\infty < x_1, x_2 < \infty.$$

We let $p_1(t) = (t, 0) \subset S_1$. We shall see that the pair $(S_1, p_1(t))$ has the Mautner property.

Example 2. Let t be a real number and let $z \in \boldsymbol{C}$, the additive group of complex numbers. Let S_2 be the ordered pairs (t, z), $t \in \boldsymbol{R}$, $z \in \boldsymbol{C}$. We define multiplication in S_2 by

$$(t_1, z_1)(t_2, z_2) = (t_1 + t_2, e^{(1-i\sigma)t_2}z_2 + z_2) \qquad \sigma > 0,\ i = \sqrt{-1}.$$

It is straightforward to verify that S_2 is a group under this law of composition. We let $p_2(t)=(t,\ 0)\subset S_2$. We shall see that the pair $(S_2,\ p_2(t))$ has the Mautner property.

LEMMA 12. *Let* $g=(x_1,\ x_2)\in S_1$, *and let* $\phi_n(g)=p_1(-n)gp_1(n)$. *Then*

$$\lim_{n\to\infty} \phi_n(g) = (x_1,\ 0).$$

Proof. $\phi_n(x_1,\ x_2)=(-n,\ 0)(x_1,\ x_2)(n,\ 0)=(x_1,\ e^{-n}x_2)$, hence $lim_{n\to\infty}\phi_n(x_1,\ x_2)=(x_1,\ 0)$.

COROLLARY 13. $lim_{n\to\infty}\phi_n(0,\ x_2)=e$, *where* e *is the identity element in* S_1. *Exactly the same statements hold for* S_2 *and* $p_2(t)$.

To be precise in S_2 we have

$$\lim_{n\to\infty} (-n,\ 0)(t,\ z)(n,\ 0) = (t,\ 0) \quad \text{or}$$

$$\lim_{n\to\infty} (-n,\ 0)(0,\ z)(n,\ 0) = (0,\ 0) = e \text{ in } S_2.$$

We let H_1 be the subgroup of S_1 of elements $(0,\ x_2)$ and H_2 be the subgroup of S_2 of elements $(0,\ z)$. *Note that* S_i *is generated by the subgroups* $p_i(t)$ *and* H_i $(i=1,\ 2)$.

THEOREM 14. *The pairs* $(S_1,\ p_1(t))$ *and* $(S_2,\ p_2(t))$ *exhibit the Mautner phenomenon.*

Proof. Let U be a unitary representation of S_i such that

$$U_{p_i(t)}\psi = \psi \qquad \text{all } t.$$

If we can show that

$$U_{h_i}\psi = \psi$$

all $h_i\in H_i$, we shall have proved that ψ is invariant under the group S_i generated by $p_i(t)$ and H_i. So let $h_i\in H_i$. Then (changing our notation slightly by setting $U_g=U(g)$):

$$\begin{aligned}(U(h_i)\psi,\ \psi) &= (U(p_i(n))U(p_i(-n)h_ip_i(n))U(p_i(-n))\psi,\ \psi)\\ &= (U(\phi_n(h_i))U(p_i(-n))\psi,\ U(p_i(-n))\psi)\\ &= (U(\phi_n(h_i))\psi,\ \psi).\end{aligned}$$

Taking the limit as n goes to infinity gives, because U is continuous,

$$(U_{h_i}\psi,\ \psi) = (\psi,\ \psi).$$

Since U_{h_i} is unitary, the length of $U_{h_i}\psi$ is that of ψ, but the above relation then implies that $U_{h_i}\psi=\psi$. This all holds for $i=1,\ 2$, which proves our theorem.

8. Proof of main theorem. Before presenting the main result of Part II Theorem 15, we shall need the following algebraic lemma.

LEMMA 15. *Let A be an automorphism of N_3 such that*

$$A = \exp \circ A^* \circ \log,$$

where the eigenvalues of A^* are e^{λ}, $e^{-\lambda}$, 1. Let $q_i(s)$ be the one parameter groups defined in Lemma 7. Then the groups

$$S_1^* = A(t)\cdot q_1(s) \quad \textit{and} \quad S_2^* = A(t)\cdot q_2(s)$$

are isomorphic to S_1, with $q_1(s)$ and $q_2(s)$ corresponding to H_1 and $A(t)$ corresponding to $P_1(t)$.

Proof. Lemma 7 tells us that

$$A(t)(q_1(s)) = q_1(e^{\lambda t}s)$$
$$A(t)(q_2(s)) = q_2(e^{-\lambda t}s).$$

But then multiplication $A(t)\cdot q_1(s)$ is given by $(t_1, s_1)(t_2, s_2) = (t_1+t_2, e^{-\lambda t_2}s_1+s_2)$ and so agrees with multiplication in S_1. Since the mapping

$$A(t)\cdot q_1(s) \to A(-t)q_2(s)$$

is an isomorphism, our assertion is verified.

Let us return to our group $A(t)\cdot N_3$ and recall certain facts.

1. $A(t)\cdot N_3$ acts as a group of measure preserving transformations on $(A(t)\cdot N_3/\Gamma, dx\times\mu)$ by right multiplication, where $A(t)\cdot N_3/\Gamma$ is the left coset space.

2. If $H = L^2(A(t)\cdot N_3/\Gamma, dx\times\mu)$, the above action induces a unitary representation U: $A(t)\cdot N_3 \to U(H)$ where $U(H)$ denotes unitary operators on H.

3. If $\psi \in H$ and $U_g\psi = \psi$ for all $g \in A(t)\cdot N_3$, then $\psi = \text{constant}$ a.e. This follows because $A(t)\cdot N_3$ acts transitively on $A(t)\cdot N_3/\Gamma$.

THEOREM 15. *Let A be an automorphism of N_3 which maps L onto itself such that A on $T_e(N_3)$ is given by a matrix*

$$\begin{bmatrix} e^{\lambda} & 0 & 0 \\ 0 & e^{-\lambda} & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \lambda \neq 0.$$

Then A_L is ergodic on N_3/L.

Proof. By Theorem 11 we see it suffices to prove that $A(t)$ acts ergodically on $(A(t)\cdot N_3/\Gamma, dx\times\mu)$. It is this last assertion that we shall actually prove. Let $H = L^2(A(t)\cdot N_3/\Gamma, dx\times\mu)$ and let $\psi \in H$ satisfy

$$U_{A(t)}\psi = \psi \qquad \text{all } t.$$

But by Statement 2 in this section the above unitary representation of $A(t)$ can be extended to the whole group $A(t)\cdot N_3$. Since S_1^* and S_2^* are subgroups of $A(t)\cdot N_3$, we get representations of S_1^* and S_2^*. But the pairs $(S_1^*, A(t))$ and $(S_2^*, A(t))$ exhibit the Mautner phenomenon, so

$$U_g\psi = \psi, \qquad g \in S_1^* \quad \text{or} \quad S_2^*.$$

Then $U_g\psi = \psi$ for all g in the group generated by S_1^* and S_2^*. But the group generated by S_1^* and S_2^* is $A(t)\cdot N_3$, so we have

$$U_g\psi = \psi, \qquad \text{all } g \in A(t)\cdot N_3.$$

But since $A(t)\cdot N_3$ acts transitively on $A(t)\cdot N_3/\Gamma$, we have $\psi=$constant a.e. This proves $A(t)$ acts ergodically on $(A(t)\cdot N_3/\Gamma,\ dx\times\mu)$, and so our assertion follows.

Part III. Affine motions on the torus

9. Torus automorphisms revisited. Before beginning the study of affine motions on the torus, let us re-examine the ergodic properties of the automorphisms of the torus using the same method of skew products and Mautner phenomena that we used on N_3/L.

Let A be an automorphism of n dimensional real vector space V^n. Then we may formulate the Jordan canonical form theorem for matrices as follows: We may write

$$V^n = V_1 \oplus \cdots \oplus V_k$$

where $AV_i = V_i$, $i=1, \cdots, k$, and let $A\,|\,V_i = A_i$. Then $A_i = D_iR_iN_i$, where all matrices on the right hand side commute, and

$$D_i = \begin{bmatrix} e^{\lambda_i} & & 0 \\ & \ddots & \\ 0 & & e^{\lambda_i} \end{bmatrix},$$

$$R_i = \begin{bmatrix} \begin{bmatrix} \cos\theta_i & \sin\theta_i \\ -\sin\theta_i & \cos\theta_i \end{bmatrix} & & \\ & \ddots & \\ & & \begin{bmatrix} \cos\theta_i & \sin\theta_i \\ -\sin\theta_i & \cos\theta_i \end{bmatrix} \end{bmatrix},$$

$$N_i = \begin{bmatrix} 1 & & * \\ & \ddots & \\ 0 & & 1 \end{bmatrix},$$

where the * denotes some matrix entries.

In general there will exist no 1-parameter group of matrices $A(t)$ such that $A(1)=A$ and $A(t)V_i=V_i$ all t. But this is merely a technical difficulty that can be overcome using complex vector spaces. Instead of introducing this extra complication, we shall merely discuss the case where $A(t)$ exists such that $A(1)=A$ and $A(t)V_i=V_i$. Thus we shall restrict ourselves to automorphisms A with the above property which we shall call property 1-P. Hence if A has property 1-P there exists a 1-parameter group $A(t)$ such that $A(t)V_i=V_i$ all t. Let $A(t)\,|\,V_i=A_i(t)$, where

$$A_i(t) = D_i(t)R_i(t)N_i(t),$$

and where all matrices in the right hand side commute for all values of t and

$$D_i(t) = \begin{bmatrix} e^{\lambda_i t} & & 0 \\ & \ddots & \\ 0 & & e^{\lambda_i t} \end{bmatrix}$$

$$R_i(t) = \begin{bmatrix} \begin{bmatrix} \cos\theta_i t & \sin\theta_i t \\ -\sin\theta_i t & \cos\theta_i t \end{bmatrix} & & \\ & \ddots & \\ & & \begin{bmatrix} \cos\theta_i t & \sin\theta_i t \\ -\sin\theta_i t & \cos\theta_i t \end{bmatrix} \end{bmatrix}$$

$$N_i(t) = \begin{bmatrix} 1 & * \\ 0 & 1 \end{bmatrix}.$$

We now proceed exactly as in Section 6.

Step 1. Form the semidirect product $A(t)\cdot V^n$, where $V^n \supset L$ and V^n/L is a torus and $A(1)$ maps L onto itself.

Step 2. Verify that if m denotes an integer, then $A(m)\cdot L=\Gamma$ is a discrete subgroup of $A(t)\cdot V^m$ and $A(t)\cdot V^m/\Gamma$ is compact.

Step 3. $A(t)$ acting on $A(t)\cdot V^m/\Gamma$ is equivalent to the skew product of A acting on V^m/Γ.

Step 4. A acts ergodically on V^m/Γ if and only if $A(t)$ acts ergodically on $A(t)\cdot V^m/\Gamma$.

Wé are now in a position to begin a Mautner phenomenon type of argument.

In our indexing of the invariant spaces V_i let us assume that for $i=1,\cdots,s$, $s\leqq k$, we have that $\lambda_i\neq 0$ in D_i.

LEMMA 16. *The subgroup $A(t)\cdot V_i$ has a subgroup S^* containing $A(t)$, isomorphic to one of the groups S_1 or S_2, and in which $A(t)$ corresponds to $p_1(t)$ or $p_2(t)$ of Section 7.*

Proof. If in $R_i(t)$, $\theta_i=0$, then $A_i(t)$ has a one-dimensional invariant subspace, call it $W_1 \subset V_i$. Then the group $A(t)\cdot W_1$ is isomorphic to S_1 exactly as in Lemma 15.

If in $R_i(t)$, $\theta_i \neq 0$, then $A_i(t)$ has a 2-dimensional invariant subspace, $W_2 \subset V_i$, and the group $A(t)\cdot W_2$ may be seen to be isomorphic to S_2 by an argument analogous to that in Lemma 15.

Thus the pair $(S^*, A(t))$ exhibits the Mautner phenomenon by Theorem 14. Now let $H = L^2(A(t)\cdot V^2/\Gamma)$, and let $\psi \in H$ satisfy

$$U_{A(t)}\psi = \psi \qquad \text{for all } t.$$

Let $H_1 \subset H$ satisfy

$$U_{A(t)}\psi = \psi \qquad \text{for all } \psi \in H_1,$$

in other words H_1 is the subspace of H where $U_{A(t)}$ acts trivially. By the Mautner phenomenon, we may conclude that

$$U_{S^*}\psi = \psi \qquad \text{all } \psi \in H_1.$$

Now by considering $A(t)\cdot V_i/S^*$ and using induction on the dimension of V_i, we see that $A(t)\cdot V_i$ acts trivially on the elements of H_1.

Another inductive argument can be used to extend easily the above to all $g \in A(t)\cdot(V_1 \oplus \cdots \oplus V_s)$. Consider the group generated by L and $V_1 \oplus \cdots \oplus V_s$, and let C be the closure of this group in V^n. Then we may apply Theorem 3.5 of [3] to conclude the following.

THEOREM 17. *The flow $A(t)$ on $A(t)\cdot V^n/\Gamma$ is ergodic if and only if the flow $A(t)$ on $A(t)\cdot V^n/C$ is ergodic.*

Thus if $s=k$ we have proved our theorem, or even if $C=V^n$ we have proved that A is ergodic on V^n/L.

We need an algebraic lemma to relate the above result to the theorems in Part I.

LEMMA 18. *Let all notation be as above. Then $C \neq V^n$ if and only if A has no eigenvalues which are roots of unity.*

Proof. If A has roots of unity as eigenvalues, then $s<k$. Let α be chosen so A^α has an eigenvalue equal to one. Then the range of $(A^\alpha - I)$ contains $V_1 \oplus \cdots \oplus V_s$, and also contains a lattice. Hence C is contained in the range of $(A^k - I)$.

If $C \neq V^n$, then $s<k$ and V^n/C_0, where C_0 is the identity component of C, contains a lattice invariant under A restricted to V^n/C_0. But all eigenvalues of A have absolute value 1, so the semisimple part of A, call it t, has the property that t^β, for some integer β, preserves a lattice in V^n/C_0. Thus t^β generates a discrete, hence finite, subgroup of a compact group. Hence $t^\gamma = e$ for some γ and t consists of roots of unity. This proves our assertion.

Thus we have a new proof of Theorem 3. Theorem 1 stays the same and is not altered by our discussion.

10. Main theorem. Again let V^n/L be a torus. Let A be an automorphism of V^n/L and let $\xi \in V^n/L$. Question: When is the transformation

$$f(x) \to A(x) + \xi$$

ergodic? This was answered in [5]. We present another proof.

We, of course, may consider the mapping of the torus onto itself given by $f' = -y \circ f \circ y(x)$, $y \in V^n/L$. Clearly f is ergodic if and only if f' is ergodic. Since

$$f'(x) = A(x) + (A - 1)y + \xi,$$

we may choose y so that ξ is not in the range of $(A-1)y$. Hence using the notation of Section 9 we see that ξ is not in V_i for any i where θ_i or λ_i are not zero. But then we may apply Theorem 7 to see that f' is ergodic on V^n/L if and only if it induces an ergodic motion on V^n/C_0. Let f^* be the image of f' on V^n/C_0. We may reason as in Section 9 and conclude that the semisimple part of f^* must act trivially in V^n/C_0. But then

$$A(t) \cdot V^n/C_0\Gamma$$

is a nil-manifold, and we may apply the results of [4] on nil-flows to conclude the following result:

THEOREM 19. *An affine transformation $f = A(x) + \xi$ on a torus is ergodic if and only if*

1. *A has no eigenvalues which are roots of unity.*
2. *f induces an ergodic transformation on $V^n/L + (A-I)V^n$, where $(A-1)V^n$ denotes the range of $A-I$.*

References

1. L. Auslander, The Mautner phenomenon and geometric flow theory, to appear.

2. ———, Modifications of solvmanifolds and G-induced flows, Amer. J. Math., 88 (1966) 615–625.

3. ———, and L. Green, G-induced flows, Amer. J. Math., 88 (1966) 43–60.

4. ———, F. Hahn and L. Green, Flows on homogeneous spaces, Ann. Math. Studies, 53, Princeton University Press, Princeton, New Jersey, 1963.

5. F. J. Hahn, On affine transformations of compact abelian groups, Amer. J. Math. 85 (1963) 428–446.

6. A. G. Postnikov, Ergodic problems in the theory of congruences and of diophantine approximations, Proc. Steklov Inst. of Math., 82 (1966), Trans. AMS (1967).

MATHEMATICS AND THINKING MATHEMATICALLY

M. L. CARTWRIGHT, Claremont Graduate School

This year I find myself in the Division of Applied Mathematics at Brown University. Not every University or College has a separate department of applied mathematics, but for over thirty years I was classed as a pure mathematician in the University of Cambridge, England, and some applied mathematicians there prefer to call themselves theoretical physicists. Moreover, I once heard a geophysicist with a mainly mathematical training say that he used 'applied mathematics' as a term of abuse, meaning stuff which was not good mathematics and not really relevant to any physical problem. All these factors have made me think about the borderline between mathematics and its applications, not only to physical problems of a more or less traditional type, but also to statistical, economic, and industrial problems.

It is well known that the origins of some of the most abstract pure mathematics can be traced through the theory of Fourier series to a problem about vibrating strings, or through the theory of irrational numbers to Greek geometry and Egyptian devices for measurement of right angles, but the pure mathematicians of the last 100 or 150 years have been pursuing the mathematics for its own sake without any thought of vibrating strings. On the other hand many major new developments in pure mathematics were initiated quite specifically for the purpose of using them in some application. For instance this is certainly true of Newton's contributions to the calculus and of probability theory, and this still seems to be happening in operational research and control theory. In distinguishing pure mathematics from applied two questions seem to arise. Is the work truly abstract and separated from all applications? And is it any more mathematical if it is truly abstract and pursued strictly for its own sake?

If we delve into the beginnings of mathematical thought in very young children or primitive peoples, there is plenty of evidence to show that the power of complete abstraction comes very slowly, and indeed to many people it probably only ever comes in a very restricted sense. A number of eminent people take

---

Mary Cartwright earned her Oxford PhD under G. H. Hardy and E. C. Titchmarsh. Since, she has been on the staff of Girton College of the University of Cambridge, serving as Director of Studies in Mathematics and then as Mistress of Girton College, holding the position of Reader in the Theory of Functions in the University of Cambridge.

Her extensive research publication in function theory includes the book *Integral Functions* Cambridge Univ. Press, 1956, 1962). She is a Fellow of the Royal Society and has received honorary degrees from Cambridge, Edinburgh, Leeds, Hull, Wales, and Oxford. She was awarded the Sylvester Medal of the Royal Society in 1964 and the De Morgan Medal of the London Mathematical Society in 1968 and was made a Dame Commander of the British Empire in June 1969 for services to mathematics.

Miss Cartwright served three years on the Council of the Royal Society, was Vice-President of the London Mathematical Society, and President in 1961–63. She also served as President of the Mathematical Association (the British counterpart to the MAA). *Editor.*

the view that thought begins with the idea of actions performed in the mind only, that is to say operations. According to Piaget an ordinary child by the time he is two can work out how he is going to do something before he does it, *provided* that the situation is simple and is familiar to him, but in order to understand abstract mathematical concepts such as 1, 2, 3, 4, · · ·, the child has to move from perceptions arising from his environment and actions to the abstractions, a long and gradual process. Much work has been done by Piaget and Innhelder on the child's conception of space, and, for instance, its powers to distinguish between different kinds of figures such as a circle, a square, and a circle with a little one either inside or outside. Their experiments have thrown much light on the development of numerical, spatial, and physical concepts of a very elementary kind among young children, but it seems doubtful to me whether the abilities tested are always truly mathematical. For young blackbirds will gape at a piece of black cardboard consisting of one large circle and two small ones attached to it, but they only gape at the small circle whose size is a certain proportion of that of the large circle. This indicates that the ability to distinguish between certain shapes may have psychological foundations.

H. and H. A. Frankfort in an essay on myth and reality point out that ancient man could reason and work out the causes of things, but worked on very different hypotheses from ours. The primitive mind asks 'who' when it looks for a cause, and cannot withdraw far from perceptual reality. When the river does not rise, the river has *refused* to rise, and so the river or the gods intend to convey something to the people. At the same time primitive man used symbols much as we do, but he can no more conceive them as signifying, yet separate from, the gods or powers than he can consider a relationship—such as resemblance—as connecting, and yet separate from, the objects compared. Hence there is a coalescence of the symbol and what it signifies, as there is coalescence of two objects compared so that one may stand for the other.

Frankfort then gives an example of this coalescence in which pottery bowls with the names of hostile tribes were solemnly smashed at a ritual by the Egyptians in the belief that real harm was done to the enemies by the destruction of their names. It may seem a far cry from this to modern mathematics, but Bochner has drawn a parallel between mathematics and myth, and replaced myth by mathematics in some of Frankfort's sentences. I am not prepared to go as far as he does by replacing the word myth by mathematics in a sentence which then asserts that mathematics transcends reasoning in that it wants to bring about the truth it proclaims. However, in the ritual we have two fundamental features of mathematics, symbols representing something and operations on those symbols representing operations on the thing itself. Symbols and notation are part of the essential basis of mathematics, and I believe that the development and standardization of a good notation is an extremely important part of the development of mathematics.

If we turn to the extreme other end of the scale, we run into another kind of difficulty in separating the mathematics from its applications. Some pure

mathematicians seem to do their mathematical thinking in terms of idealized physical and spatial ideas. The late G. H. Hardy, who taught me, was very much against applied mathematics, but in a footnote to a joint paper with J. E. Littlewood published in a Swedish periodical he wrote that a certain problem is most easily grasped in terms of cricket averages. Norbert Wiener would translate a mathematical problem into the language of Brownian motion, and I believe that his thinking was completely abstract although I do not know the theory, or remember what he said well enough to be quite sure. Hadamard has described his visualization of the proof that there is a prime greater than 11. To consider all prime numbers from 2 to 11, i.e., 2, 3, 5, 7, 11 he visualized a confused mass. Forming the product $2\times3\times5\times7\times11=N$, since N is large, he visualized a point remote from the mass. Increasing the product by 1 he saw another point a little beyond the first. $N+1$, if not a prime, is divisible by a prime greater than 11; Hadamard saw a place between the mass and the first point. This seems to me to be a sort of mathematical shorthand and would certainly have to be translated back to numbers before it could be communicated to anyone else.

As I said earlier, I have until now always been classed as a pure mathematician, but Professor J. E. Littlewood and I did a lot of work on the theory of ordinary differential equations arising from problems of radio engineering. Littlewood is also a very pure mathematician in many ways, but he worked on antiaircraft gun fire in the First World War, and he translated our problems, which were suggested by radio values and oscillations, capacitance and inductance, etc., into dynamical problems and called all the solutions of our equations 'trajectories' as if they were the paths of missiles shot from a gun. In the radio problems there are oscillations with negative damping, and so we had periodic trajectories going up and down over and over again, and I am sure that the abstraction was complete although there was often a certain woolliness until the argument was complete, just as in Hadamard's visualization. Between these two extremes there are some users of complicated mathematics, physicists and engineers in particular, who are thinking all, or nearly all, the time in terms of the physics of the problem. Engineers have consulted me about a number of different types of problem, radio, control theory, oscillations of stretched wires; they usually come with some equations and very little explanation. I have to ask a lot of questions before they tell me everything relevant to the mathematical problem. It seems difficult for them to think in abstract mathematical terms, the symbols to them seem to mean the engineering concepts, currents and circuit constants such as impedance and inductance. This is important in two ways. The engineers have mental reservations and can check at every stage because they visualize how the physical system works. On the other hand they find it difficult to apply the mathematical processes used in one field to any other physical problem, even if they are just as relevant there. Some years ago at a conference for engineers I was asked to speak on Liapunov's method for stability problems. I described the basic principles as simply as I could, and after I spoke

Professor Parks lectured on applications of the method. Many in the audience commented that the order of our lectures should have been reversed, and that they would have understood my lecture much better if they had understood that I was talking about the phase plane. It is possible that this was partly a question of notation and terminology, but I believe that they could do advanced mathematics best by thinking of it in terms of their particular engineering problems. The Liapunov method was developed mainly in connection with control engineering and by now has adopted much of its terminology, but the mathematics arising there need to be abstracted and put in a form which makes it available in connection with other applications. Problems of ordinary differential equations have arisen in connection with astronomy, ballistics, radio engineering, control theory, mechanical oscillations of machinery; each application has special features, and the theory of it was often developed in a correct logical form quite a long way before it was fitted into the general theory of ordinary differential equations as pure mathematics. The individual who formulated the equation and asked the question is, in the sense of my title, thinking mathematically, but he is not doing mathematics until he operates on his symbols. Please note that I do not say 'asked for a solution of the equation' because, although he may say that, he really wants to know something about the solutions in general. Is there a periodic solution? Is it stable? Will it remain stable if I change a certain parameter? Will the period be longer or shorter? He may find the methods which he needs in the literature and do the work himself. He may find a mathematician to help him. Although I myself have helped to develop the general theory and settle certain theoretical problems, I do not think that I have ever produced a result useful for any specific practical problem when it was needed. For soon after Littlewood and I began work on these problems, it was realized that the variations in individual thermionic valves was so great that precise mathematical results were not worth the trouble, and satisfactory experimental determinations could be more easily obtained. In recent times the person who formulates the mathematical statement of a physical or other real life problem usually does not do anything very original in the mathematical handling of it, although some interesting purely mathematical work on matrices appears in journals concerned with computing or applications to economics, detached from other pure mathematics.

To sum up so far I believe that the dividing line between strictly abstract thinking in mathematics and thinking in terms of the real world is by no means clearly defined and some of the major developments in mathematics such as the Calculus were thought out more or less in terms of the real world. Further abstraction does not necessarily make the mathematics any better. For the Babylonian schoolmasters constructed sets of most complicated artificial formulae, perhaps 200 on one tablet, for their pupils to simplify. Their mathematics was sufficiently abstract for them to be indifferent whether they added the number of men to the number of days. In present circumstances this seems abstraction at its worst, but perhaps then it was a step forward. The Babylonians must

have developed the laws of arithmetic a long way to set these complicated exercises, but mainly for practical purposes whether it was accounting or astronomy.

Now let us turn to those who do mathematics for its own sake. I should like to begin with the Hindu who in about 1200 B.C. wrote, "As crests on the heads of peacocks, as the gems on the hoods of snakes, so is ganita, mathematics, at the top of the sciences known as the Vedanga." Ganita is literally the science of calculation and in the early days it consisted of finger arithmetic, mental arithmetic, and higher arithmetic in general. At first it included astronomy, but geometry belonged elsewhere. At one stage higher mathematics was called 'dust work' because it was done in sand spread on the board or on the ground. We owe our so-called Arabic numerals to the Hindus, and they advanced a long way in algebra very early.

Most people consider that the Greeks were the first to do mathematics for its own sake and to realize the need for proof. The word 'mathema' meant originally a subject of instruction, but very early it was restricted to mathematical subjects among which Pythagoras included geometry, theory of numbers, sphaeric (or spherical trigonometry used for astronomy), and music. They classified numbers not only as odd and even, but as even-even, 2^m; even-odd, $2(2n+1)$; odd-even $2^{m+1}(2n+1)$, and also proved that there are an infinity of primes. I doubt whether they could calculate as well as the Babylonians, but probably that did not attract them, and also they lacked the incentives provided by the government of a far flung empire. I feel that I have to remind myself of the difficulties due to the absence of convenient symbols. Sir Thomas Heath writing of the arithmetic of Nicomachus said 'If the verbiage is eliminated, the mathematical content can be stated in quite a small compass,' but Heath used modern notation and Arabic numerals. In the Wasps of Aristophanes one of the characters tells his father to do an easy sum 'not with pebbles but with fingers,' and Herodotus says that, in reckoning with pebbles, Greeks move left to right, Egyptians right to left, which implies vertical columns facing the reckoner.

The Greeks also developed a theory of geometry which remained more important than any other for nearly 2,000 years, and was the first deliberate development of a logical system in mathematics. In the third century A.D. an unknown writer jokingly used words of Homer intended for something else to describe mathematics:

Small at her birth, but rising every hour.
She stalks on earth and shakes the world around.

For, says Anatolius, Bishop of Laodacia, who quoted it, mathematics begins with a point and a line and forthwith it takes in the heaven itself and all things within its compass. If this was the Greek viewpoint at such a late date, is it possible that their geometry was not truly abstract and that the symbols of point and line were still partly coalesced with the abstract point and line?

The position of geometry and more generally spatial concepts in mathematics is not completely clear to me. In recent times all types of geometry have been

given an analytical basis and freed from the logical difficulties such as those which used to worry schoolmasters teaching about congruent triangles by the method of superposition. I therefore ask myself whether geometry and spatial concepts are really part of the basis of mathematics or a field of application similar to mechanics, both terrestrial and celestial, or to games of chance. The reason for the traditional special position of geometry may be that in geometry the symbols are the objects themselves; the abstract point, line, and triangle are represented by a point, line, and triangle; what is more, so long as the geometry is plane geometry they can be drawn on a flat surface by pen or pencil on paper or in sand on the ground. When Greek geometry was being developed there was no good notation for dealing with numbers, and even in the 15th Century the solution of a cubic equation was described in geometrical terms and illustrated by a figure for lack of a good algebraic notation. In mechanics a comparable real life representation of motion could not be used to explain the theory; written symbols or geometrical figures were needed for communication. But if we ask whether the contributions of spatial concepts to modern mathematics are greater than those of other real life problems it is difficult to answer. Spatial thinking has led to the highly abstract theory of irrational numbers of Cantor and Dedekind, and permeates mathematical thought in almost all fields; the physical sciences have given rise to the calculus (not without the help of geometry), and statistics and probability have their basis in multitudinous practical problems.

Pfeiffer explains the situation well in relation to probability. Some of the salient points in his account are as follows: The history of probability theory (as is true of most theories) is marked both by brilliant intuition and discovery and by confusion and controversy. Until certain patterns had emerged to form the basis of a clear-cut theoretical model, investigators could not formulate problems with precision, and reason about them with mathematical assurance.

From what some people say it sounds to me as if quantum theory had not yet reached this stage, but it is certainly beyond my competence to form a valid judgment.

Pfeiffer continues by saying that although long experience was needed to produce a satisfactory theory, we need not retrace and relive the fumblings which delayed the discovery of an appropriate mathematical model. That is a mathematical system whose concepts and relationships correspond to the appropriate concepts and relationships of the real world. Once the model has been discovered, studied, and refined, it becomes possible for an ordinary mind to grasp, in a reasonably short time, a pattern which took decades of effort and the insight of genius to develop in the first place. I note that Pfeiffer asserts that the most successful model of probability theory known at present is characterized by considerable mathematical abstractness.

J. Willard Gibbs wrote 'One of the principal objects of theoretical research in any department of knowledge is to find the point of view from which the subject appears in its greatest simplicity' and Bushaw says that one of the dis-

tinctive characteristics of modern mathematics is its way of taking old mathematical ideas apart like watches, studying the parts separately, and putting these parts together again in new and interesting combinations and studying these complications in turn. I believe that this process has contributed enormously to this simplification in mathematics itself, and so made it more readily available for applications. Mandelbrojt referring to the quotation from Willard Gibbs says 'Integration in function spaces provided such a point of view over and over again in widely scattered areas of knowledge and it gave us not only a new way of looking at problems but actually a new way of thinking about them.' Now one might call Fréchet, the father of abstract spaces, and in the front of his book he puts a quotation from Hadamard's survey of functional analysis given in 1911. 'The functional continuum does not present any simple concept to our imagination. Geometrical intuition tells us nothing *a priori* about it. We are forced to remedy this ignorance and we can do it only analytically, by creating a chapter of the theory of sets for handling the functional continuum.' Elsewhere Hadamard wrote that the calculus of variations was nothing but the first chapter of functional analysis, and of his own work on the calculus of variations, hyperbolic partial differential equations, and certain other topics he said that he owed the greater part to his contacts with the physicist Duhem, through Duhem's book on hydrodynamics, elasticity, and acoustics and many conversations when they were both at Bordeaux. So we have a record here of the complete cycle from a physical basis through the calculus of variations to functional analysis and abstract spaces, and thence to a multitude of applications through the process of analyzing geometrical ideas and putting them together again in a most abstract new way to create function spaces.

A further variation on this pattern has become evident of recent years and that is the use of an auxiliary model consisting of various graphical, mechanical, and other aids to visualizing, remembering, and even discovering things about the mathematical model. The visual images of Hadamard, Hardy's cricket averages, and Littlewood's trajectories might be considered as auxiliary models, but of more universal significance are the analogue machines with electronic devices which simulate what happens in, for instance, fluid mechanics, or rather what corresponds in the mathematical model. We now have

(A) The real world of actual phenomena, known to us by various ways of experiencing these phenomena.

(B) The abstract world of the mathematical model which uses symbols to state relationships and facts with great precision and economy.

(C) The auxiliary model.

The transition from A to B is the formulation of real world phenomena in mathematical terms; the transition B to A is the interpretation of the deduction by pure mathematics from that formulation. Both these I consider to be thinking mathematically, but only the deductions inside B are mathematics. We may also think mathematically by moving from B to C which is a secondary inter-

pretation, and then either back to B to confirm what C has suggested or from C direct to A.

As Pfeiffer points out, the value of both the mathematical model and the auxiliary model depends on how successfully the appropriate features of the model may be related to the 'real-life' situation. The models cannot be used to prove anything about the real world, although a study of it may help us to discover important facts about the real world. A model is not true or false; it fits or it does not fit. It is unsatisfactory if either (1) the solutions of the model problems have unrealistic interpretations, for instance, arbitrarily large quantities or arbitrarily fine differences, or (2) it is incomplete or inconsistent so that the mathematics produces contradictions. Many models fit amazingly well. Karl Pearson wrote 'The mathematician, carried along on his flood of symbols, dealing apparently with purely formal truths, may still reach results of endless importance for our description of the physical universe.'

Until perhaps 100 years ago many scientists and mathematicians knew a bit of everything, and the mathematical formulation, as I said of Newton in particular, was done by someone who was a good enough mathematician to develop the mathematics to a considerable extent. This is particularly true of Sir Isaac Newton, but in these days of specialization the scientist or economist, or worker in close contact with the real world situation must do stage A→B. Sir Cyril Hinshelwood, former President of the Royal Society, said 'Scientists need to be taught mathematics as a language they can actually speak. It is of great importance for the scientist to be able to learn the art of formulating problems in mathematical terms which of course is a quite difficult job. You have to think very accurately and carefully about a problem before you can do it. You have to have practice in speaking the language of mathematics. It does not matter being an expert in differential equations. You can go to the expert for help in solving an equation. But you cannot expect the mathematician to do the translation into mathematics. There should be an early and rather intensive cultivation of the power of thinking about real things and the application of mathematical symbolism to physical ideas.' He went on to draw a parallel between learning simple French as a child and learning to express physical ideas in mathematics when the level of physics and mathematics reached are both elementary, so that the child becomes accustomed to the process by easy stages. Although he advocates, as I do, that the scientist should do the mathematical formulation, his words seem to imply an incomplete abstraction. In his mind the mathematical symbols were still representing their physical counterparts, not that this matters for a scientist who has access to an expert mathematician, but it is clear from Mandelbrojt's remarks on function spaces and Hadamard's remarks about the functional continuum that without complete abstraction on the part of some mathematicians we should lack some of the most expressive parts of the mathematical language used by scientists.

* * *

Since delivering the above lecture I came across the following passage by A. Robinson in *Some thoughts on the history of mathematics*, Compositio Math., 20 (1968) 188–193.

'Euclid's geometry was supposed to deal with real objects, whether in the physical world or in some ideal world. The definitions which preface several books in the Elements are supposed to communicate what object the author is talking about even though, like the famous definition of the point and the line, they may not be required in the sequel. The fundamental importance of the advent of non-Euclidean geometry is that by contradicting the axiom of parallels it denied the uniqueness of geometrical concepts and hence, their reality. By the end of the nineteenth century, the interpretation of the basic concepts of geometry had become irrelevant. This was the more important since geometry had been regarded for a long time as the ultimate foundation of all mathematics. However, it is likely that the independent development of the foundations of the number system which was sparked by the intricacies of analysis would have deprived geometry of its predominant position anyhow.'

Although it confirms my views on Euclidean geometry, it does not seem to recognize the geometrical origin of the theory of irrational numbers.

I also noticed that A. Aaboe in *Episodes from the Early History of Mathematics*, writes 'Even the oft repeated statement that the Egyptians knew the 3, 4, 5 right angle has no basis in available tests, but was invented about 80 years ago.'

The above lecture was given as the Samuel Newton Taylor Lecture at Goucher College, January 30, 1969.

The author is a Visiting Professor at Brown University, supported by the United States Air Force under Grant No. AF-AFOSR 693-67 and by the United States Army under Contract No. DA-31-124-ARO-D-270 and by the National Science Foundation under Contract No. GP-9024.

References

1. S. Bochner, The role of mathematics in the rise of science, Princeton, 1966.
2. D. Bushaw, Elements of General Topology, Wiley, New York, 1963.
3. B. Datta and A. N. Singh, A History of Hindu Mathematics, Lahore, 1935.
4. H. and H. A. Frankfort, The intellectual adventures of ancient man, Chicago, 1946.
5. T. L. Heath, A history of Greek mathematics, Vol. 2, Oxford, 1921.
6. S. Mandelbrojt, Les tauberiens généraux de Norbert Wiener, Bull. Amer. Math. Soc., 72 (1966) 48–51.
7. O. Neugebauer, The exact sciences in antiquity, Acta Hist. Sci. Nat. Medicinalium, Copenhagen, 9 (1951).
8. P. E. Pfeiffer, Concepts of Probability Theory, McGraw-Hill, New York, 1965.
9. J. Piaget, B. Innhelder, and A. Sjeminska, A child's conception of geometry, Trans. by E. A. Lunzer, Basic Books, New York, 1960.
10. N. Tinbergen, The Herring Gull's World, Basic Books, New York, 1961.

SHARPENED FORMS OF THE PLANE ISOPERIMETRIC INEQUALITY

DONALD C. BENSON, University of California, Davis

1. Introduction. By *the* isoperimetric inequality we mean here the following: *Let* Γ *be a simple closed rectifiable curve in* E^2 *with perimeter* L *and area* A. *Then*

$$L^2 - 4\pi A \geqq 0, \tag{1.1}$$

and equality holds only if Γ *is a circle.*

By introducing parameters of the curve other than L and A, it is possible to obtain sharpened forms of (1.1) in which the right hand side is replaced by a nonnegative quantity which is not identically zero. See [1, 80–83, 111–113] and [8, p. 591] for examples of this sort. (In [1, p. 81] certain formulas are incorrect; see [2, p. 140].)

In Section 2, a sharpened form of (1.1) is found which yields in special cases two different inequalities of T. Kubota [3]. (The methods used are quite different from those of Kubota.)

In Section 3, it is assumed that the curve Γ is contained in an annulus with outer and inner radii r_1 and r_2, respectively. It is further assumed that Γ winds once around the inner circle and that it contains points both on the inner and outer circles. A sharpening of (1.1) is given in which r_1 and r_2 occur as parameters.

In considerations of this sort it is usual to use the theory of convex sets. This will not be done here; the results are all demonstrated directly for rectifiable curves. The methods are analytical rather than geometrical. In each case one expresses the geometrical quantities by certain integrals. The results are obtained by using Cauchy's inequality and by transforming some of the integrals by a suitable substitution.

The author has obtained inequalities by similar methods in [4] and [5].

2. A generalization of two inequalities of T. Kubota. Let Γ be a simple closed rectifiable curve in the x-y plane. Let w be the width of Γ in some arbitrary direction. Choose a coordinate system so that Γ lies in the strip $|y| \leqq w/2$ and so that Γ contains points on the lines $y=w/2$ and $y=-w/2$.

We parametrize Γ with the positive orientation by means of arc length s. It is known that x and y on Γ are absolutely continuous functions of s on Γ. (See [6, p. 17].) Putting x' and y' for dx/ds and dy/ds, and observing that x' and y' exist for almost all s on Γ and satisfy $x'^2+y'^2=1$, we have the following expression for the total arc length L of Γ:

$$L = \int_\Gamma ds = \int_\Gamma (x'^2 + y'^2)^{1/2} ds. \tag{2.1}$$

From Cauchy's inequality we have

$$(2.2)\qquad \begin{aligned}(x'^2+y'^2)^{1/2} &= [(2y/w_0)^2-(1-(2y/w_0)^2)]^{1/2}(x'^2+y'^2)^{1/2}\\ &\geqq (|2y|/w_0)|x'| + (1-(2y/w_0)^2)^{1/2}|y'|,\end{aligned}$$

where w_0 is a parameter which must satisfy $w \leqq w_0$ so that the square root in (2.2) is real on Γ. We shall discuss special choices for w_0 later. Integrating (2.2) we have

$$(2.3)\qquad L \geqq \frac{2}{w_0}\int_\Gamma |yx'|\,ds + \int_\Gamma \left(1-\left(\frac{2y}{w_0}\right)^2\right)^{1/2} |y'|\,ds.$$

Equality holds in (2.3) if and only if the first order differential equation is

$$(2.4)\qquad (w_0^2-(2y)^2)^{1/2}|x'| = 2|yy'|$$

satisfied almost everywhere. A closed curve Γ is a solution of (2.4) if and only if it can be formed by fitting together arcs of circles of radius $w_0/2$ centered on the x-axis and segments of the lines $y=w_0/2$ and $y=-w_0/2$. (E.g., see figure 1.)

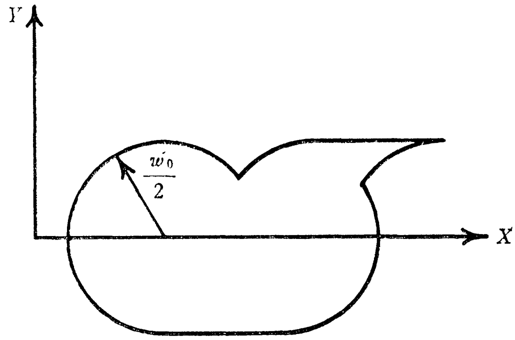


FIG. 1

For the area A enclosed by Γ we have

$$(2.5)\qquad A = \left|\int_\Gamma y\,dx\right| = \left|\int_\Gamma yx'\,ds\right| \leqq \int |yx'|\,ds$$

with equality if and only if yx' has the same algebraic sign at all points at which x' is defined. We may choose arc length so that $y(0)=-w/2$ and for $s=s_1$ we have $y(s_1)=w/2$. We treat the second term on the right of (2.3) as follows, using a standard theorem on integration by substitution [7, p. 221]:

$$(2.6)\qquad \begin{aligned}\int_\Gamma \left(1-\left(\frac{2y}{w_0}\right)^2\right)^{1/2}|y'|\,ds &\geqq \int_0^{s_1}\left(1-\left(\frac{2y}{w_0}\right)^2\right)^{1/2} y'ds\\ &\quad + \int_{s_1}^{L}\left(1-\left(\frac{2y}{w_0}\right)^2\right)^{1/2}(-y')ds\\ &= 2\int_{-w/2}^{w/2}\left(1-\left(\frac{2y}{w_0}\right)^2\right)^{1/2} dy = \frac{w}{w_0}(w_0^2-w^2)^{1/2} + w_0 \arcsin\frac{w}{w_0}\cdot\end{aligned}$$

Equality holds in (2.6) if and only if y' is nonnegative on $[0, s_1]$ and nonpositive on $[s_1, L]$. Putting together (2.6), (2.5), and (2.3) we have

$$L \geqq \frac{2A}{w_0} + \frac{w}{w_0}(w_0^2 - w^2)^{1/2} + w_0 \arcsin \frac{w}{w_0}, \tag{2.7}$$

for any $w_0 \geqq w$. Inequality (2.7) is the main result of this section.

Now let us look more carefully at the conditions under which equality holds in (2.7). We may assume, without loss of generality, that Γ is parametrized by arc length $s(0 \leqq s \leqq L)$ with the usual positive orientation, and that we have, as above, $y(0) = y(L) = -w/2$, $y(s_1) = w/2$. Then the conditions mentioned following (2.3), (2.5), and (2.6) give the following necessary and sufficient conditions for equality in (2.7):

(a) The curve Γ consists of arcs of circles with radius $w_0/2$ centered on the x-axis and possibly segments of the lines $y = w_0/2$ and $y = -w_0/2$. (Note that the line segments can occur only if $w = w_0$.)

(b) On Γ, x is an increasing function of s wherever y is positive, and decreasing wherever y is negative. (The positive orientation of Γ is used here.)

(c) On Γ, y is nondecreasing for s in $[0, s_1]$ and non-increasing for s in $[s_1, L]$.

The curves Γ which satisfy (a), (b), and (c) fall into two different types, depending on whether equality holds in (2.7) with a choice of w_0 such that $w_0 < w$ or such that $w_0 = w$. If equality holds with $w_0 < w$, then (a), (b), and (c) imply that Γ is a "lens" of diameter w, i.e., the boundary of the intersection of two circular discs with centers on the x-axis and diameter w_0. If equality holds with $w_0 = w$, then (a), (b), and (c) imply that Γ is a "racetrack," i.e., the boundary of the convex hull of two circular discs situated as above; in this case note that w is the *minimum width* of Γ. (Of course, we intend that a circle of radius w is a special case both of the racetrack and the lens.)

Suppose the width w in an arbitrary direction and the area A of a closed curve Γ are given. How shall we choose w_0 in order to give the best estimate of L in (2.7)?

If $A \geqq \pi w^2/4$, then there is a racetrack having A as its area and w as its minimum width. Putting $w = w_0$, equality holds in (2.7) for this racetrack. This shows that the right hand side of (2.7) is maximum for this choice of w_0, and (2.7) reads

$$L \geqq 2A/w + \pi w/2. \tag{2.8}$$

REMARK. If it is known that w is actually the diameter d of Γ, then the foregoing case occurs only if Γ is a circle of diameter d because the inequality of Bieberbach ([**1**] p. 76) asserts

$$A \leqq \pi d^2/4$$

with equality only in case Γ is a circle of diameter d.

Inequality (2.8) was observed by Kubota [**3**]. The isoperimetric inequality

(1.1) follows from it because (2.8) may be written in the form

$$L^2 - 4\pi A \geqq \left(\frac{2A}{w} - \frac{\pi w}{2}\right)^2. \tag{2.9}$$

If $A < \pi w^2/4$, then (2.8) holds, but a sharper result is possible. In this case there is a lens with area A and diameter w. The right hand side of (2.7) will be maximized if we put for w_0 the diameter of the circular discs whose intersection gives this lens. Putting $w/w_0 = \sin\phi$, we see that to satisfy this geometrical condition we must choose w_0 so that the transcendental equation

$$2A/w^2 = \phi \csc^2 \phi - \cot \phi \tag{2.10}$$

is satisfied. One can show that if $A < \pi w^2/4$, then (2.8) has exactly one solution. Thus for $A < \pi w^2/4$ (2.7) gives us

$$L \geqq 2w\phi^* \csc \phi^* \tag{2.11}$$

where $\phi = \phi^*$ is the solution of (2.10).

We have shown that (2.8) and (2.11) give sharp estimates for the length of Γ if the width w in an arbitrary direction and the area A of a closed curve Γ are given. We may also consider the question of determining the best possible upper bound for A if w and L are fixed. Inequality (2.7) may be written

$$2A \leqq w_0 L - w(w_0^2 - w^2)^{1/2} - w_0^2 \arcsin (w/w_0). \tag{2.12}$$

For what choice of w_0 is the right hand side of (2.12) the best possible upper bound for the area A? There are two cases to consider. Either we have $L \geqq \pi w$ or $2w \leqq L < \pi w$. (We always have $2w \leqq L$.) In the first case ($L \geqq \pi w$) there exists a racetrack with the length L and minimum width w which gives equality in (2.12) with $w = w_0$. Thus (2.8) is the best lower bound for A if L and w are fixed and satisfy $L \geqq \pi w$. However, the case $2w \leqq L < \pi w$ gives something different. As before, put $\sin \phi = w/w_0$. Here we find that there is a lens with length L and diameter w which gives equality in (2.12) provided ϕ satisfies the transcendental equation

$$2\phi w = L \sin \phi, \tag{2.13}$$

and, with ϕ^{**} equal to the unique solution of (2.13), (2.12) reads

$$A \leqq \frac{L}{8\phi^{**}} (L - 2w \cos \phi^{**}). \tag{2.14}$$

Inequality (2.14) is due to Kubota in the same paper [3] which contains inequality (2.8).

REMARK. If the area, the diameter d, and the minimum width w^* of Γ are known, one may consider the problem of maximizing the right hand side of (2.7) as a function of w and w_0 subject to the constraints $0 \leqq w_0 \leqq w$ and $w^* \leqq w \leqq d$. It does not seem of interest, however, to do so, because the lower bound

for L which results is not sharp, i.e., in general there will not exist a curve Γ with prescribed A, d, and w^* which gives equality in (2.7). To give such a sharp lower bound for L appears to be an interesting problem.

3. An inequality involving L, A, r_1, and r_2. Again let Γ denote a simple closed rectifiable curve in E^2. Let the origin for a polar coordinate system be a point 0 inside the curve.

Let us choose arc length s for parameter of Γ in such a way that s gives the positive orientation to Γ. Then since x and y on Γ are absolutely continuous functions of s, it follows that the polar coordinates, r and θ, are absolutely continuous functions of Γ. (See [6, p. 17].) The derivatives dr/ds and $d\theta/ds$ are defined for almost all s on Γ.

Let r_1 and r_2 denote, respectively, $\sup\{r \mid (r, \theta)\in\Gamma\}$ and $\inf\{r \mid (r, \theta)\in\Gamma\}$. Let E and F denote positive constants such that the relations

$$\begin{aligned} Fr_1 - Er_1^{-1} &= 1 \\ Fr_2 - Er_2^{-1} &= -1 \end{aligned} \tag{3.1}$$

hold unless $r_1 = r_2$, in which case Γ is a circle with center at 0. In fact, it is not difficult to compute explicitly from (3.1) the following

$$\begin{aligned} r_1 &= (2F)^{-1}(1 + (1 + 4EF)^{1/2}) \\ r_2 &= (2F)^{-1}(-1 + (1 + 4EF)^{1/2}), \end{aligned} \tag{3.2}$$

$$\begin{aligned} F &= (r_1 - r_2)^{-1} \\ E &= r_1r_2(r_1 + r_2)(r_1 - r_2)^{-2}. \end{aligned} \tag{3.3}$$

From Cauchy's inequality we have, putting r' and θ' for dr/ds and $d\theta/ds$, respectively,

$$(r'^2 + r^2\theta'^2)^{1/2}(1 + 4EF)^{1/2} \geqq (Fr + Er^{-1})r\,|\,\theta'\,| + (1 - (Fr - Er^{-1})^2)^{1/2}\,|\,r'\,|. \tag{3.4}$$

Equality holds in (3.4) if and only if the differential equation

$$(1 - (Fr - Er^{-1})^2)^{1/2} r\,|\,\theta'\,| = (Fr + Er^{-1})\,|\,r'\,| \tag{3.5}$$

holds a.e. Evidently one solution to (3.5) is

$$\sin(\theta + c) = Fr - Er^{-1}, \tag{3.6}$$

which is the equation of a circle with $\sup r = r_1$ and $\inf r = r_2$. Other solutions are the circles $r = r_1$ and $r = r_2$. Let us put L for the length of Γ and A for the area enclosed by Γ. Noticing $r'^2 + r^2\theta'^2 = 1$ a.e., and using a standard theorem on integration by substitution we have

$$\begin{aligned} (1 + 4EF)^{1/2}L &\geqq \int (Fr + Er^{-1})r\,|\,\theta'\,|\,ds + \int (1 - (Fr - Er^{-1})^2)^{1/2}\,|\,r'\,|\,ds \\ &\geqq 2FA + 2\pi E + \int (1 - (Fr - Er^{-1})^2)^{1/2}\,|\,r'\,|\,ds. \end{aligned} \tag{3.7}$$

Equality holds in the first inequality of (3.7) if and only if the differential equation (3.5) holds a.e. Equality holds in the second inequality of (3.7) if and only if $\theta' \geqq 0$ a.e. Now, transforming the integral on the right hand side of (3.7) by means of the substitution $u = Fr - Er^{-1}$, we have

$$
\begin{aligned}
&\int_{\Gamma} (1 - (Fr - Er^{-1})^2)^{1/2} | r' | \, ds \\
&\qquad \geqq \int_{r=r_2}^{r=r_1} (1 - (Fr - Er^{-1})^2)^{1/2} r' ds - \int_{r=r_{\cdot}}^{r=r_2} (1 - (Fr - Er^{-1})^2)^{1/2} r' ds \\
(3.8) \qquad &\qquad = \frac{1}{F} \int_{-1}^{1} (1 + u(u^2 + 4EF)^{-1/2})(1 - u^2)^{1/2} du \\
&\qquad = \frac{1}{F} \int_{-1}^{1} (1 - u^2)^{1/2} du = \pi/2F,
\end{aligned}
$$

using (3.1) and noting that $u(u^2+4EF)^{-1/2}(1-u^2)^{1/2}$ is an odd function. Equality holds in (3.8) if and only if Γ consists of two disjoint connected arcs Γ_1 and Γ_2 such that r is a nondecreasing function of s on Γ_1 and nonincreasing on Γ_2. Putting (3.7) and (3.8) together we have the inequality

$$
(3.9) \qquad L \geqq (2FA + 2\pi E + \pi/2F)(1 + 4EF)^{-1/2}.
$$

The necessary and sufficient conditions for equality in (3.9) are just the union of the conditions for equality in (3.7) and (3.8). One sees that equality holds in (3.9) only for the circles (3.6).

Inequality (3.9) is the goal of this section; it is a sharpened form of the isoperimetric inequality (1.1). This can be seen by rewriting (3.9) in the form

$$
(3.10) \qquad L^2 - 4\pi A \geqq \frac{(2FA - 2\pi E - \pi/2F)^2}{1 + 4EF}.
$$

This research was supported in part by National Science Foundation Grant GP-8185.

References

1. T. Bonnesen and W. Fenchel, Theorie der konvexen Körper, Chelsea reprint, New York, 1948.

2. M. Sholander, On certain minimum problems in the theory of convex curves, TAMS, 73 (1952) 139–173.

3. T. Kubota, Einige Ungleichheitsbeziehungen über Eilinien und Eiflächen, The Science Reports of the Tôhoku Imperial University, Ser. 1, 12 (1923) 45–65.

4. D. C. Benson, Elementary solutions of variational problems, J. of Math. and Mech., 17 (1968) 1103–1116.

5. ———, Inequalities involving integrals of functions and their derivatives, J. Math. Anal. and Applications, 17 (1967) 292–308.

6. Lamberto Cesari, Surface Area, Princeton University Press, Princeton, 1956.

7. L. M. Graves, The Theory of Functions of Real Variables, Second edition, McGraw-Hill, New York, 1956.

8. H. Flanders, A proof of Minkowski's inequality for convex curves, this MONTHLY, 75 (1968) 581–593.

ULAM'S CONJECTURE AND GRAPH RECONSTRUCTIONS

P. V. O'NEIL, College of William and Mary

1. Introduction. Henri Poincaré once meditated upon how it can happen that some people do not understand mathematics. Perhaps one reason is that most worthwhile questions nowadays are too difficult to explain. Anyone in a position to understand the problem can usually follow the solution. In this connection, Ulam's conjecture is a pleasant example of a mathematical rarity—a nontrivial, easily stated problem that anyone can take a crack at. At least in its graph-theoretic formulation, Ulam's conjecture is simpler to explain than, say, baseball or Monopoly. Also, reminiscent of seventeenth and eighteenth century mathematics, the clever amateur probably stands as good a chance as anyone of settling it.

In this paper, we discuss Ulam's conjecture as it relates to graph theory, together with some results and questions in the area of graph reconstructions to which it naturally leads. Theorems and proofs will be stated very informally, and, with the exception of the existence theorem of Section 5, none are new.

2. Ulam's version. In his famous problem book [**10**, p. 29], Ulam asked the following question:

Suppose A and B are sets with n elements each ($n \geqq 3$). A metric p is given on A with the property that $p(x, y)$ is either 1 or 2 whenever x and y are in A and $x \neq y$. A similar metric is given on B. Now suppose that the $n-1$ element subsets of A and B can be labelled, $A_1, \cdots, A_n$ and $B_1, \cdots, B_n$, in such a way that each A_i is isometric to B_i. Does this force A to be isometric to B?

Ulam's conjecture is that it does.

This certainly seems a plausible guess, and this author has even been told that it is obvious, although not by anyone who has attempted to write down a proof. We shall not be concerned directly with the statement just given, but shall instead turn immediately to a graph theoretic version. The advantages to be gained are concreteness and a certain visual appeal which the term isometry lacks by itself. These may be more than just fanciful, as most of the progress in this area seems to have been made by adopting the graph theorist's point of view.

3. A graph theoretic formulation. Surprisingly, the barest essentials of graph theory are sufficient to restate the problem. We shall briefly review these below. The interested reader can pursue the subject in [**1, 3, 4, 8, 9**].

A *graph* may be thought of as a finite, nonempty set of points (vertices) and lines (branches) connecting distinct points. It is easy to show that any graph

Prof. O'Neil earned his Rensselaer PhD under P. Slepian in 1965. He held an instructorship at the Univ. of Minnesota before his present position at William and Mary. His research is in graph theory. *Editor.*

can be drawn in 3-space so that branches intersect only at vertices (for a proof, see [3, p. 8]). For obvious reasons, graphs are in practice usually drawn in the plane. Personal taste is allowed, but disregarded, in the drawing of graphs. Thus, the graphs of Figure 1 are considered the same, because a bit of stretching and bending (no tearing) will make them identical. Such graphs are said to be *isomorphic.* More carefully, G and K are isomorphic if there is a 1-1 function f between their vertices such that a branch between x and y is in G exactly when the branch between $f(x)$ and $f(y)$ is in K.

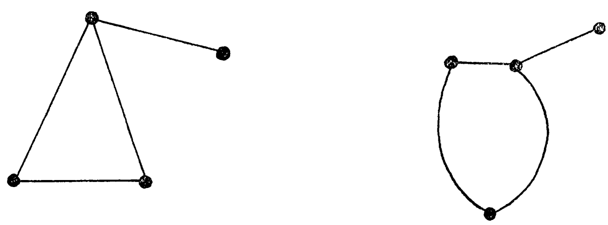


FIG. 1

Finally, if x is a vertex of G, then $G-x$ denotes the graph obtained by deleting x and all branches incident with x from G.

We now have enough to represent Ulam's conjecture graphically. Given the elements $a_1, \cdots, a_n$ of A, mark n points $x_1, \cdots, x_n$ as vertices of a graph G_A. Interpret $p(a_i, a_i)=0$ to mean there is no branch from x_i to x_i. If $i\neq j$, draw a branch between x_i and x_j exactly when $p(a_i, a_j)=1$. Similarly, draw a graph G_B to represent the metric space (B, q). Ulam's conjecture now reads:

If G_A-x_i is isomorphic to G_B-y_i for each i, then G_A is isomorphic to G_B.

A slightly different point of view is provided by the notion of a graph reconstruction. Suppose we start with n graphs, $H_1, \cdots, H_n$, each with $n-1$ vertices. A graph K is called a *reconstruction* of the H_i's if K has n vertices, $t_1, \cdots, t_n$, such that $K-t_i$ is essentially (isomorphic to) H_i. In the above discussion, G_A is a reconstruction of the graphs G_A-x_i, as is G_B of the G_B-y_i. If we regard G_A-x_i as essentially the same graph as G_B-y_i (through the isomorphism), and call it, say, H_i, then we have the following version of the problem.

If K and K' are reconstructions of $H_1, \cdots, H_n$, then K and K' are isomorphic.

Thus, the problem becomes one of essential uniqueness of graph reconstructions. This is Ulam's conjecture in its most concrete form.

4. Some results on graph reconstructions. There are two quite different ways of looking at graph reconstructions. First, given a reconstruction G of $G-x_1, \cdots, G-x_n$, are there any essentially different reconstructions? Of equal interest is a second point of view: given n graphs $H_1, \cdots, H_n$, each with $n-1$ vertices, does a reconstruction exist?

The reader can convince himself that the graphs of Figure 2 have no recon-

struction (certain vertices are labelled for later reference). Thus, the existence given in Ulam's conjecture is by no means a trivial assumption. We shall see in Section 5 exactly how nontrivial it is. In this section, we shall discuss some results which have been achieved concerning uniqueness and construction of reconstructions, assuming existence.

The question of uniqueness has been settled for "small" graphs. A theorem due to Kelly [7] says that Ulam's conjecture is true if n is less than 7. This has been extended to $n=7$ by Harary and Palmer [5]. Unfortunately, the arguments used (consideration of all cases) do not suggest a method of proof for arbitrary n.

Some of the best results to date depend upon connectivity conditions. G is said to be *connected* if any two distinct vertices of G can be linked by a path made up of branches of G. In Figure 2, H_1 is connected while H_5 is not. Here H_5 has four *components* (maximal connected subgraphs), and H_4, two components. A *tree* is a connected graph with no circuits (H_1 and H_2 are trees, while H_3, H_4 and H_5 are not).

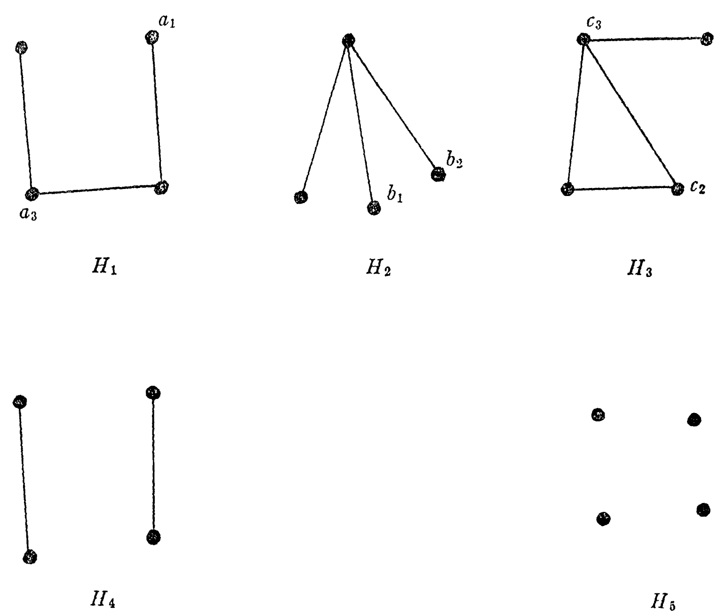
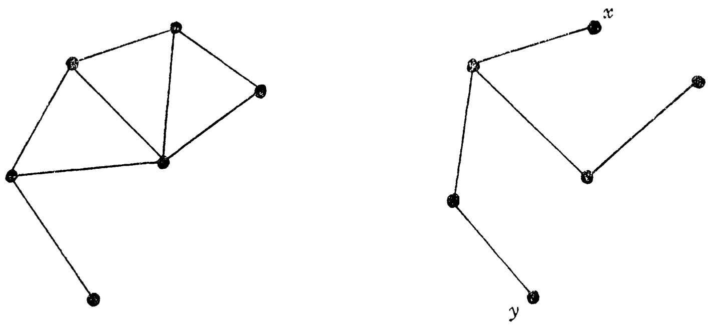


FIG. 2

In the paper referred to above, Kelly was able to show that any tree is uniquely reconstructible from its $n-1$ vertex subgraphs. Thus, Ulam's conjecture is true for trees. The reader might try translating the assumption that G is a tree back into the original language of the problem. Harary and Palmer [6] improved upon Kelly's theorem by showing that any tree can in fact be uniquely reconstructed from just the subgraphs obtained by removing the *endpoints*

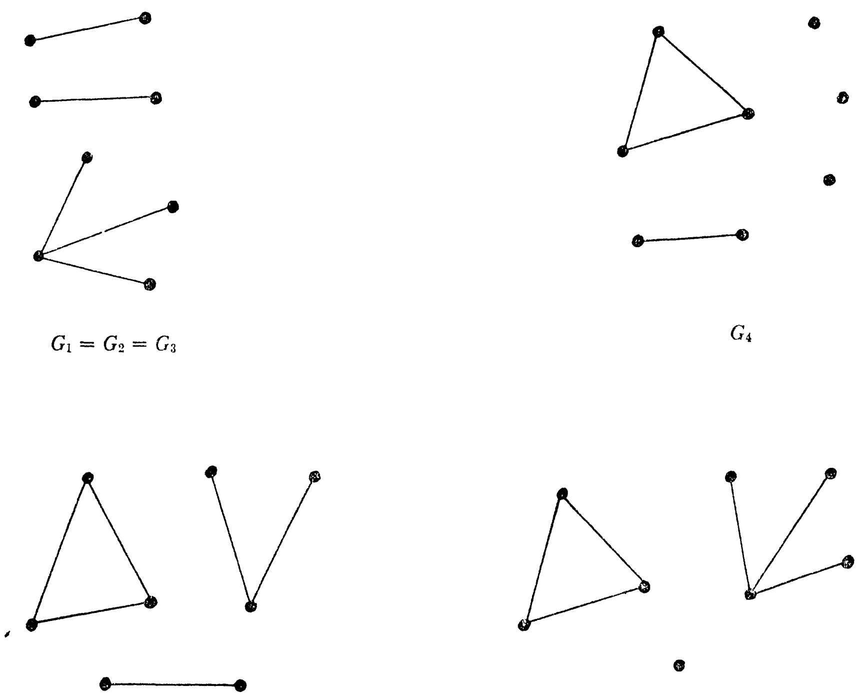
(vertices with just one incident branch). A more recent improvement, which we shall not discuss here, can be found in [2].

The following result is quite remarkable. Suppose the graphs $G_1, \cdots, G_n$,



G T

FIG. 3



$G_5 = G_6 = G_7$, $G_8 = G_9$

FIG. 4

each with $n-1$ vertices, have a reconstruction G. If at most one G_i is connected, then G is unique.

The hypothesis on the number of connected G_i appears surprising at first. In fact, it is equivalent to saying that any reconstruction which exists must be disconnected. For suppose G is connected. It is easy to see that some set of $n-1$ branches of G forms a tree T (Figure 3). Further, since $n>2$, T has at least two end points, say x and y. Then, at least $G-x$ and $G-y$ are connected. It is straightforward to show that, conversely, if G is disconnected, then all but possibly one G_i are disconnected.

The disconnectedness of any reconstruction can be exploited to actually produce one, component by component. For simplicity, the constructions will be given for particular graphs. The reader might try extending them to the general case.

First consider G_1 through G_9 as shown in Figure 4. Here, all G_i are disconnected. The trick is to observe that the smallest number (greater than 1) of components in any G_i is 3, and to try to produce a suitable G with this number of components. Of those G_i with three components, choose one, say G_9, with a smallest component. For convenience, label its components C_1, C_2 and C_3, with C_3 the smallest (see Figure 5). Keep C_1 and C_2 as candidates for components of

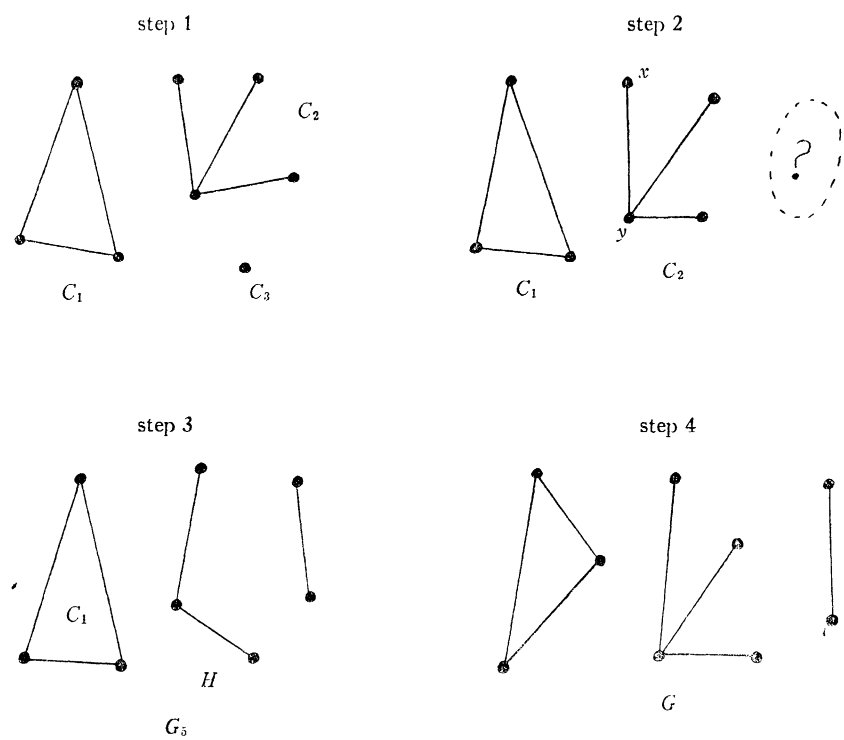


FIG. 5

G. There remains to produce one more. Take C_1 or C_2, say C_2, and remove a vertex which does not disconnect C_2, say x (but not y), forming a graph H. Now identify some G_i having three components, two being C_1 and H. A good choice is G_5. The third component of G_5 can now be taken as the final component of G.

Note that G as constructed has no *isolates* (vertices with no incident branches). This could have been predicted from the given G_i: G will have k isolates exactly when k of the G_i have $k-1$ isolates and each of the remaining $n-k$ has at least (but possibly more than) k isolates. In this event, the construction just given fails. However, when G has isolates, it is particularly easy to reconstruct. For example, in Figure 6, we have $k=3$, and we can simply pick some G_i (say G_1) having two isolates and adjoin another one.

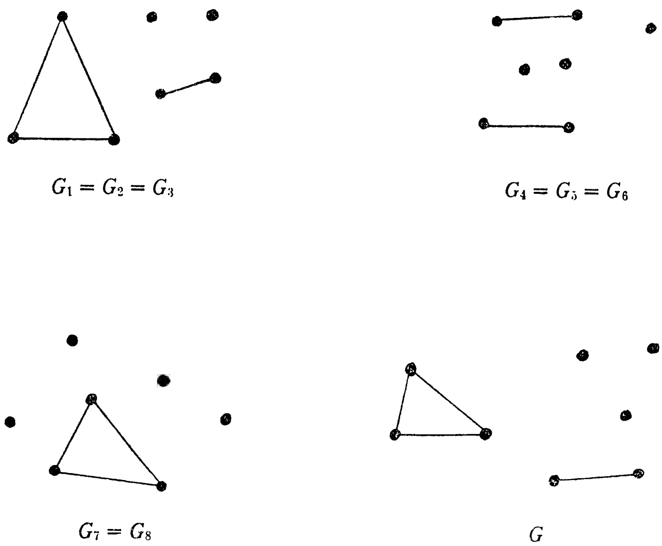


FIG. 6

To sum up, if a reconstruction G of the graphs G_i exists, we can: (1) predict from the G_i when G must be disconnected, and (2) in this event, construct G uniquely.

These results produce another immediately upon taking complements. The *complement* of G, denoted G^c, is formed by erasing all branches of G and drawing in any which were not originally there. It is easy to see that G is a reconstruction of $G_1, \cdots, G_n$ exactly when G^c is a reconstruction of $G_1^c, \cdots, G_n^c$. In Figure 5, note that G^c is connected, so this is an honest extension to some graphs not already treated.

The constructions we have just described are special cases of the general ones discussed by Harary [4], who also has additional details and references on graph reconstructions.

5. A more general reconstruction problem. We have already seen (Figure 2) that n graphs may very well have no reconstruction. To see how bad the situation can be, note that even H_3 and H_5 by themselves have none; that is, there is no five vertex graph G such that $G-x$ is H_3 and $G-y$ is H_5. In this section, we focus attention on the general existence problem: given $G_1, \cdots, G_k$, each with $n-1$ vertices ($k \leqq n$), does there exist an n vertex G with vertices $x_1, \cdots, x_k$ such that $G-x_i$ is essentially G_i? Ideally, we would like necessary and sufficient conditions on the G_i's which will tell us when to expect a reconstruction, and then a method for finding it (or them).

To get some feeling for what is happening, we begin with the simplest nontrivial case, $k=2$. A first impression is that G_1 and G_2 have to be "almost the same" for there to be a reconstruction. Specifically, if $G-x_1 = G_1$ and $G-x_2 = G_2$, then

$$(G - x_1) - x_2 = G_1 - x_2$$

and

$$(G - x_2) - x_1 = G_2 - x_1.$$

But it is easy to see that $(G-x_1)-x_2$ and $(G-x_2)-x_1$ are identical. Thus, G_1 and G_2 are the same "to within one vertex." More carefully: if a reconstruction of G_1 and G_2 exists, then there must be vertices x of G_1 and y of G_2 such that G_1-x is isomorphic to G_2-y.

This is a very reasonable necessary condition in that it is both easily stated and applied (try it on H_3 and H_5 of Figure 2). Happily, it is also sufficient. For suppose $G_1-x=G_2-y$. We can produce a suitable G by starting with a copy of G_1-x and drawing in two new vertices and appropriate branches to each. In Figure 2, we have $H_1-a_1=H_3-c_2$, and two possible constructions are indicated in Figure 7, (a) and (b), where $G-u=H_1$ and $G-v=H_3$.

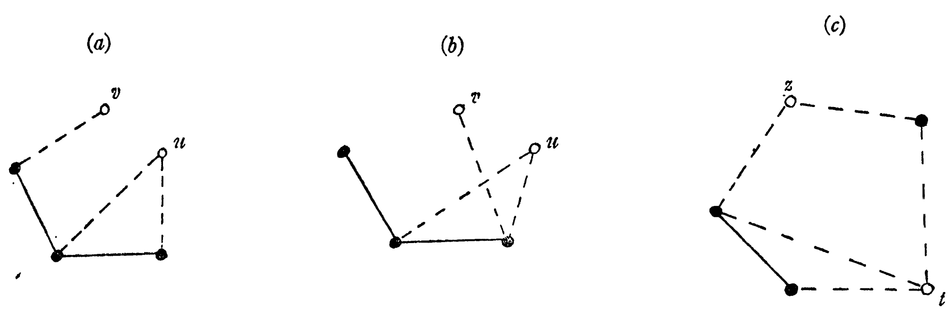


FIG. 7

Incidentally, this shows that a reconstruction is never unique when $k=2$, since we are free to put in or leave out the branch between the two adjoined vertices (u and v of Figure 7 (a) and (b)). It also suggests a method for producing

all possible reconstructions of two graphs. Simply list all pairs of vertices a and b such that $G_1-a=G_2-b$, and repeat the constructions for each pair. For example, referring again to Figure 2, note that $H_1-a_3=H_3-c_3$. A reconstruction K begun with a copy of H_1-a_3 is given in Figure 7 (c), where $K-t=H_1$ and $K-z=H_3$.

Having found a nice solution for $k=2$, we are led naturally to attempt a similar theorem for the general case. Half of the above reasoning carries through immediately. If $G-x_i=G_i$, then

$$(G - x_i) - x_j = G_i - x_j = (G - x_j) - x_i$$
$$= G_j - x_i,$$

so that, in pairs, G_i and G_j are identical to within a vertex. This isn't surprising, since any reconstruction of k graphs is a reconstruction of any two of them. The bubble bursts, however, with H_1, H_2 and H_3 of Figure 2. In pairs, H_1-a_1 and H_2-b_1; H_2-b_2 and H_3-c_2; and H_3-c_3 and H_1-a_3, are isomorphic. There is not even any cheating, since no vertex has been used twice. But there is no reconstruction of H_1, H_2 and H_3.

It turns out that, to insure existence, we need not only isomorphisms between G_i-x and G_j-y, but particular kinds of isomorphisms. This is best explained in terms of vertex orderings. If x_i^j is an ordering of the vertices of G_j, for $1\leqq i\leqq n-1$ and $1\leqq j\leqq k$, then we can define maps f_{1+r}^t from $G_t-x_r^t$ to $G_{1+r}-x_t^{1+r}$ by:

$$x_i^t \to x_i^{1+r} \qquad \text{if } 1 \leqq i \leqq t-1 \text{ or } 1+r \leqq i \leqq n-1,$$

and

$$x_i^t \to x_{1+i}^{1+r} \qquad \text{if } t \leqq i < r.$$

Here, values of t are specified by $1\leqq t\leqq \min(3, k-1)$, and, for each t, values of r by $t\leqq r\leqq k-1$. Now, in general, these maps may have no particular properties of interest. In the event that they are all isomorphisms, we call the orderings *compatible*. A general existence theorem can now be stated.

THEOREM. *The graphs $G_1, \cdots, G_k$ have a reconstruction if and only if their vertices can be ordered compatibly.*

Most of the details of the proof are uninteresting, with the exception of the constructive part. When the vertex orderings are compatible, a reconstruction can be produced as follows. Mark n points, say $v_1, \cdots, v_n$, as vertices. Draw a branch between v_i and v_j for $2\leqq i<j\leqq n$ exactly when x_{i-1}^1 and x_{j-1}^1 are connected by a branch in G_1. For $i\geqq 3$, connect v_1 with v_i when x_1^2 and x_{i-1}^2 are connected in G_2. Finally, if $k>2$, connect v_1 and v_2 if x_1^3 and x_2^3 are connected in G_3.

Two observations are in order. First, G is constructed from G_1, G_2 and G_3. The compatibility conditions simply insure that the other G_i are recoverable from G. Second, the conditions are more theoretical than practical, as there is no clear cut way of determining whether or not there are compatible orderings.

It happens that, when $k=2$, the vertices of G_1 and G_2 can always be ordered so that the isomorphism $G_1-x\leftrightarrow G_2-y$ is of the required form. For $k>2$, however, complications mount rapidly. Unfortunately, at least for the present, there does not seem to be any way to avoid this in the general case.

To bring the discussion full cycle, it is easy to verify that the following is an equivalent statement of Ulam's conjecture.

If x_i^j and y_i^j, where $1\leqq i\leqq n-1$, are compatible vertex orderings of G_j, for $1\leqq j\leqq n$, then, for each j, the map $x_i^j\rightarrow y_i^j$ is an isomorphism of G_j onto itself.

This formulation clearly lacks the simple appeal of the others. It does serve to pinpoint the real difficulty built into the conjecture, as any proof will in effect show that compatible orderings are essentially unique.

6. Conclusion. Our concluding remarks will be confined to a brief statement of the current extent of ignorance about graph reconstructions.

Virtually nothing is known, for example, about when uniqueness starts happening. For $k=2$, reconstructions are never unique. For $k=3$, uniqueness may occur. For example, three complete graphs of $n-1$ vertices each have exactly one reconstruction, namely the complete graph of n vertices (complete means that every possible branch is drawn in). When else can three graphs uniquely determine a reconstruction? When do k graphs uniquely determine a reconstruction? Is it possible to characterize a reasonable class of n vertex graphs which are uniquely reconstructible from certain subgraphs (as Kelly did for trees and all the $n-1$ vertex subgraphs, and Harary and Palmer did for trees and their maximal subtrees)?

Such questions are interesting in their own right, and even partial solutions would certainly yield some insight into the general problem of existence and uniqueness of graph reconstructions.

References

1. C. Berge, Theory of Graphs and Its Applications, Wiley, New York, 1962.

2. J. A. Bondy, On Kelly's congruence theorem for trees, Proc. Cambridge Philos. Soc., 65 (1969) 387–399.

3. R. G. Busacker and T. L. Saaty, Finite Graphs and Networks, McGraw-Hill, New York, 1965.

4. F. Harary, Graphical Reconstructions, appears in A Seminar on Graph Theory, F. Harary, ed., Holt, Rinehart and Winston, 1967, 18–20.

5. F. Harary and E. M. Palmer, On similar points of a graph, J. Math. Mech., 15 (1966) 623–630.

6. ———, The reconstruction of a tree from its maximal subtrees, Canad. J. Math., 18 (1966) 803–810.

7. P. J. Kelly, A congruence theorem for trees, Pacific J. Math., 7 (1957) 961–968.

8. O. Ore, Graphs and Their Uses, Random House, New York, 1963.

9. ———, Theory of Graphs, AMS Publications, Vol. 38, Providence, R. I., 1962.

10. S. M. Ulam, A Collection of Mathematical Problems, Wiley (Interscience), New York 1960.

AN APPLICATION OF QUASIGROUPS TO GEOMETRY

DAVID MERRIELL, Vassar College

There exist in recent literature several applications of quasigroups to geometry. The purpose of this article is to show how identities in a particular idempotent medial quasigroup may be used to prove some theorems in elementary geometry.

A *quasigroup* is a binary system $(Q, \cdot)$ which is closed under its operation and in which for any a and b in Q, the equations $ax=b$ and $ya=b$ have solutions in Q. A binary system is *idempotent* if $a^2=a$ for every a in Q. It is *medial* if

$$(ab)(cd) = (ac)(bd) \tag{1}$$

for every a, b, c, d in Q. Different authors have called medial quasigroups abelian, entropic, symmetric, and alternation quasigroups. The algebraic properties of such systems have been studied intensively. For idempotent medial quasigroups, the following identities hold:

$$a(bc) = (ab)(ac) \quad \text{left-distributive law;} \tag{2}$$

$$(ab)c = (ac)(bc) \quad \text{right-distributive law;} \tag{3}$$

$$a(ba) = (ab)a \quad \text{flexible law.} \tag{4}$$

If $(F, +, \cdot)$ is a field and r is any element of F, then the system $(F, \triangle)$ in which $a \triangle b$ is defined to be $(1-r)a+rb$ is an idempotent medial quasigroup. Within these quasigroups, excluding the cases where $r=0$, 1, the left and right cancellation laws hold:

$$a \triangle b = a \triangle c \text{ implies } b = c, \quad \text{and} \quad a \triangle c = b \triangle c \text{ implies } a = b. \tag{5}$$

If "$\triangle$" and "$\circ$" are two operations corresponding to elements r and s in F, then

$$(a \triangle b) \circ (c \triangle d) = (a \circ c) \triangle (b \circ d). \tag{6}$$

Also, for any k in F

$$k(a \triangle b) = (ka) \triangle (kb), \tag{7}$$

and in the special case where $s=1/2$, one obtains the identities

$$(a \triangle b) + (c \triangle d) = (a + c) \triangle (b + d); \tag{8}$$

$$(a \triangle b) + (b \triangle a) = a + b; \tag{9}$$

$$a + (b \triangle c) = (a + b) \triangle (a + c). \tag{10}$$

Now let us consider an example of a medial quasigroup, given by Sholander [2], in which Q is the set of points in the Euclidean plane, α is a fixed triangle, and ab is the third vertex of the triangle β with vertices (a, b, ab) which is directly similar to α. This means that β is similar to α and has the same orientation. It seems natural to require that $a^2=a$ in this model. Thus the quasigroup is idempotent.

Let the points (x, y) of the plane be represented by complex numbers $x+iy$. Then if the fixed triangle α has vertices represented by the complex numbers $(0, 1, r)$, the vertices of β are $(a, b, a\triangle b)$ where $a\triangle b=(1-r)a+rb$. Therefore all of the above identities apply. Corresponding to each identity there is a geometric property, in most cases not immediately obvious, of directly similar triangles. We shall use the symbols a and b for the vertices as well as for the complex numbers which represent them. We shall also use capital letters $A, B, C, \cdots$ to represent the vertices corresponding to $a, b, c, \cdots$.

In [3], I. M. Yaglom proved the following theorem: *Given an arbitrary convex quadrilateral ABCD, if equilateral triangles are constructed exterior to the quadrilateral on sides AB and CD and interior to it on BC and DA, then the free vertices of the equilateral triangles form a parallelogram.* Our identity (8) may be used to prove the following generalization: Given an arbitrary quadrilateral $ABCD$, if triangles directly similar to a fixed triangle are constructed on the directed sides AB, CB, CD, and AD, then the free vertices of the triangles form a parallelogram. The free vertices are $a\triangle b$, $c\triangle b$, $c\triangle d$, and $a\triangle d$. The diagonals of the quadrilateral which they form intersect at $[(a\triangle b)+(c\triangle d)]/2$ and $[(c\triangle b)+(a\triangle d)]/2$ which are both equal to $[(a+c)\triangle(b+d)]/2$. Hence the quadrilateral is a parallelogram.

The quasigroup identities can also be used to generalize results of Garfunkel and Garfunkel [1]. Following their lead, we define three directed segments A_1B_1, A_2B_2, A_3B_3 to be *r-symmetric* if and only if, when the triangles $A_iB_iC_i$ are constructed directly similar to $(0, 1, r)$ where r is any complex number not on the real line, then the triangle $C_1C_2C_3$ is also directly similar to $(0, 1, r)$. The authors call the three segments "equi-Pappian" in case the fixed triangle is equilateral, that is, if $r=(1\pm i\sqrt{3})/2$.

In this more general context, the first three theorems of [1] become:

THEOREM 1. *If ABC and $A'B'C'$ are two triangles directly similar to* $(0, 1, r)$, *then the joins AA', BB', CC' of corresponding vertices are r-symmetric.*

COROLLARY. *Three directed line segments AA', AB', AC' are r-symmetric if and only if $A'B'C'$ is directly similar to* $(0, 1, r)$.

THEOREM 2. *Quadrilateral $ABCD$ is a parallelogram if and only if AD, BC, and AC are r-symmetric.*

THEOREM 3. *If the three sides AB, CD, and BC of a directed quadrilateral are r-symmetric, then BD, AC, and AD are r-symmetric.*

In the proofs which follow, $x\triangle y=(1-r)x+ry$. The triangle XYZ is directly similar to $(0, 1, r)$ if and only if $x\triangle y=z$.

Proof of Theorem 1. Using (1), $(a\triangle a')\triangle(b\triangle b')=(a\triangle b)\triangle(a'\triangle b')=c\triangle c'$.

Proof of Corollary. Let $(a\triangle a')\triangle(a\triangle b')=a\triangle c'$. Since $(a\triangle a')\triangle(a\triangle b')=(a\triangle a)\triangle(a'\triangle b')=a\triangle(a'\triangle b')$, we use (5) to obtain $a'\triangle b'=c'$. Hence $A'B'C'$ is directly similar to $(0, 1, r)$. The converse is a direct consequence of the theorem.

Proof of Theorem 2. If AD, BC, and AC are r-symmetric, then $(a\triangle d)\triangle(b\triangle c)=a\triangle c$. Hence $(1-r)^2a+(1-r)r(d+b)+r^2c=(1-r)a+rc$ so that $r(1-r)(d+b)=r(1-r)(a+c)$. Since $r\neq 0$, 1, $a+c=b+d$. This implies that $ABCD$ is a parallelogram. The argument is reversible. Note that the condition $(a\triangle d)\triangle(b\triangle c)=a\triangle c$ is necessary and sufficient for $ABCD$ to be a parallelogram.

Proof of Theorem 3. It is given that $(a\triangle b)\triangle(c\triangle d)=b\triangle c$. By (1), $(b\triangle d)\triangle(a\triangle c)=(b\triangle a)\triangle(d\triangle c)$.Using first (9) and then (8), together with the fact that $-(x\triangle y)=(-x)\triangle(-y)$,

$$\begin{aligned}(b \triangle a) \triangle (d \triangle c) &= [a+b-(a\triangle b)]\triangle[c+d-(c\triangle d)]\\ &= [(a+b)\triangle(c+d)]-[(a\triangle b)\triangle(c\triangle d)]\\ &= [(a\triangle d)+(b\triangle c)]-(b\triangle c)=a\triangle d.\end{aligned}$$

Therefore BD, AC, and AD are r-symmetric.

Finally, we use this method to prove Theorem 4 of [1]: If equilateral triangles are constructed outwardly on the alternate sides AB, CD, EF of a directed hexagon so that their free vertices P, Q, R form an oppositely directed equilateral triangle, then the free vertices of the equilateral triangles constructed on the other alternate sides BC, DE, FA also form an oppositely directed equilateral triangle.

Here the vertices P, Q, R are $a\triangle b$, $c\triangle d$, and $e\triangle f$ and we are given that $(c\triangle d)\triangle(a\triangle b)=e\triangle f$. A necessary and sufficient condition for the triangle $(a, b, a\triangle b)$ to be equilateral is that $(a\triangle b)\triangle a=b$. By (4), $(a\triangle b)\triangle a=a\triangle(b\triangle a)$. Therefore,

$$\begin{aligned}(c\triangle a)\triangle(d\triangle b) &= e\triangle f,\\ c\triangle a=(d\triangle b)\triangle(e\triangle f) &= (d\triangle e)\triangle(b\triangle f),\\ (b\triangle f)\triangle(c\triangle a) &= d\triangle e,\\ (b\triangle c)\triangle(f\triangle a) &= d\triangle e,\\ (d\triangle e)\triangle(b\triangle c) &= f\triangle a.\end{aligned}$$

Since $b\triangle c$, $d\triangle e$, and $f\triangle a$ are the free vertices of the second set of triangles, this completes the proof.

References

1. J. Garfunkel and S. Garfunkel, The equi-Pappian property, this MONTHLY, 74 (1967) 1217–1221.

2. M. Sholander, On the existence of the inverse operation in alternation groupoids, Bull. Amer. Math. Soc., 55 (1949) 746–757.

3. I. M. Yaglom, Geometric transformations, Random House, New York, 1962.

AN ANALOGUE OF PTOLEMY'S THEOREM IN SPHERICAL GEOMETRY

J. E. VALENTINE, Utah State University

1. Introduction. In a note [3] in this journal, L. M. Blumenthal and B. E. Gillam exhibited, among other things, the geometrical significance of the cofactors of the determinant

$$\Delta(p_1, p_2, \cdots, p_{n+2}) = \left| \cos(p_ip_j/\rho) \right|,$$

where $i, j=1, 2, \cdots, n+2$, and $p_1, p_2, \cdots, p_{n+2}$ are $n+2$ points of the n-dimensional spherical surface $S_{n,\rho}$ (the "surface" of a sphere of radius ρ in euclidean $(n+1)$-dimensional space, with shorter arc metric, where $p_ip_j=p_jp_i$ denotes the distance of the points p_i, p_j).

Haantjes [4] gave a proof of the spherical analogue of the ptolemaic inequality, and [5] developed techniques which give a new proof of Ptolemy's Theorem and its converse in the euclidean plane. It is further stated [5] that these techniques give proofs of the analogue of Ptolemy's Theorem and its converse in the spherical and hyperbolic planes. However, his analogue of the converse of Ptolemy's Theorem in the hyperbolic plane is false; e.g., the determinant $\left|\sinh^2 p_ip_j/2\right| =0$, where $i, j=1, 2, 3, 4$ for four points on a horocycle in the hyperbolic plane (see [6]), and consequently the result for the spherical plane is not immediate.

The purpose of this note is to give the geometrical significance of the determinant

$$\gamma(p_1, p_2, p_3, p_4) = \left| \sin^2 p_ip_j/2\rho \right|,$$

where $i, j=1, 2, 3, 4$, and p_1, p_2, p_3, p_4 are four points of the spherical plane of radius ρ. The determinant $\gamma(p_1, p_2, p_3, p_4)$ is a principal minor of a determinant obtained from $\Delta(p_1, p_2, p_3, p_4)$. The spherical analogue of Ptolemy's Theorem and its converse is thus obtained. Though the techniques and proofs are quite different from Haantjes's [5], the theorem and its converse in the spherical plane are the same. Moreover, the techniques used here give an immediate generalization.

The analogue of Ptolemy's Theorem and its converse is more apparent if it is remembered that the determinant

$$D(p_1, p_2, p_3, p_4) = \begin{vmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & p_1p_2^2 & p_1p_3^2 & p_1p_4^2 \\ 1 & p_1p_2^2 & 0 & p_2p_3^2 & p_2p_4^2 \\ 1 & p_1p_3^2 & p_2p_3^2 & 0 & p_3p_4^2 \\ 1 & p_1p_4^2 & p_2p_4^2 & p_3p_4^2 & 0 \end{vmatrix}$$

vanishes for each quadruple of points p_1, p_2, p_3, p_4 of the euclidean plane (see [1], pp. 153, 154).

It is noted that Ptolemy's Theorem and its converse for the euclidean plane may be stated as follows:

THEOREM (Ptolemy). *Four points p_1, p_2, p_3, p_4 of the euclidean plane lie on a circle or line if and only if the determinant*

$$C(p_1, p_2, p_3, p_4) = \left| p_ip_j^2 \right|, \quad \textit{where} \quad i, j = 1, 2, 3, 4 \quad \textit{vanishes.}$$

2. Preliminary Remarks. It is well known [2, pp. 162–163] that the determinants

$$\Delta(p_1, p_2, p_3, p_4) = \left| \cos(p_ip_j/\rho) \right|, \quad \text{where} \quad i, j = 1, 2, 3, 4,$$

and

$$\Delta(p_1, p_2, p_3, p_4, p_5) = \left| \cos(p_ip_j/\rho) \right|, \quad \text{where} \quad i, j = 1, 2, 3, 4, 5,$$

vanish for each quadruple and for each quintuple of points of the 2-dimensional spherical surface of radius ρ. Moreover, $\Delta(p_1, p_2, p_3) = \left|\cos\,(p_ip_j/\rho)\right| \geqq 0$, where $i, j = 1, 2, 3$, and vanishes if and only if the three points are on a great circle.

3. Analogue of the Ptolemaic Inequality. A metric space is said to satisfy the ptolemaic inequality provided for each quadruple p, q, r, s of its points, the three products $pq \cdot rs$, $pr \cdot qs$, $ps \cdot qr$ of "opposite" distances satisfy the triangle inequality. It is known [2, p. 80, exercise 5] that the spherical plane is not ptolemaic. The spherical plane, however, does satisfy an analogous inequality.

THEOREM 3.1. *If p_1, p_2, p_3, p_4 are four points, then the determinant*

$$\gamma(p_1, p_2, p_3, p_4) = \left| \sin^2(p_ip_j/2\rho) \right| \leqq 0,$$

where $i, j = 1, 2, 3, 4$, and $\gamma(p_1, p_2, p_3, p_4) = 0$ if p_1, p_2, p_3, p_4 lie on a great circle.

Proof. Since $\Delta(p_1, p_2, p_3, p_4) = 0$, it follows that the determinant obtained from $\Delta(p_1, p_2, p_3, p_4)$, by bordering it with a first row and a first column with common element -1, and the rest of the elements in the first row all zeroes, and the rest of the elements in the first column all ones, also vanishes.

The result of subtracting the first column of this bordered determinant from the second, third, fourth, and fifth columns, respectively, and making use of the fact that $1 - \cos x = 2\sin^2(x/2)$ is

$$(1) \qquad \begin{vmatrix} -1 & \cdot & 1 \\ & \cdot & \\ \cdots & \cdots & \cdots \\ & \cdot & \\ 1 & \cdot & -2\sin^2(p_ip_j/2\rho) \end{vmatrix} = 0, \quad \text{where} \quad i, j = 1, 2, 3, 4.$$

If three of the four points do not lie on a great circle, assume the labeling so that $\Delta(p_2, p_3, p_4) \neq 0$. Let [1, 2] denote the cofactor of the element in the

first row and second column; an expansion theorem for determinants yields:

$$-2^4\gamma(p_1, p_2, p_3, p_4)\Delta(p_2, p_3, p_4) - [1, 2]^2 = 0.$$

Since $\Delta(p_2, p_3, p_4) > 0$, it follows that $\gamma(p_1, p_2, p_3, p_4) \leqq 0$.

If each triple of the four points lies on a great circle, then the rank of $\Delta(p_1, p_2, p_3, p_4)$ is two, and consequently the rank of the determinant in (1) is three. Thus, in this case, $\gamma(p_1, p_2, p_3, p_4) = 0$.

COROLLARY. *If p, q, r, s is a quadruple of points, then the three products of the sines of half the "opposite" distances satisfy the triangle inequality.*

Proof. By Theorem 3.1, $\gamma(p, q, r, s) \leqq 0$. However, expansion of $\gamma(p, q, r, s)$ and routine but tedious computations yield $\gamma(p, q, r, s) = A\cdot B\cdot C\cdot D$, where

$$\begin{aligned} A &= -\ [\sin\lambda\sin\delta + \sin\epsilon\sin\theta + \sin\alpha\sin\beta] \\ B &= [\sin\lambda\sin\delta + \sin\epsilon\sin\theta - \sin\alpha\sin\beta] \\ C &= [\sin\lambda\sin\delta - \sin\epsilon\sin\theta + \sin\alpha\sin\beta] \\ D &= [-\sin\lambda\sin\delta + \sin\epsilon\sin\theta + \sin\alpha\sin\beta] \end{aligned}$$

and $\lambda = pq/2\rho$, $\delta = rs/2\rho$, $\epsilon = pr/2\rho$, $\theta = qs/2\rho$, $\alpha = ps/2\rho$, and $\beta = qr/2\rho$.

Since $\gamma(p, q, r, s) \leqq 0$, it follows that exactly one of B, C, D cannot be negative. If exactly one of B, C, D were negative, then $\gamma(p, q, r, s) \geqq 0$, and hence $\gamma(p, q, r, s) = 0$. Thus one of the other two factors B, C, D would be equal to zero. Addition of the negative factor and the zero factor would contradict the fact that the sum of two nonpositive numbers is nonpositive. In exactly the same manner, it follows that no two of the factors B, C, D can be negative. Therefore, B, C, D are all nonnegative.

4. Analogue of Ptolemy's Theorem. Theorem 3.1 shows that $\gamma(p, q, r, s) = 0$ if p, q, r, s are points on a great circle. We now establish the following:

THEOREM 4.1. *If p, q, r, s are four points on a circle, then $\gamma(p, q, r, s) = 0$.*

Proof. If p, q, r, s are points on a circle, $C(c{:}pc)$, with center c and radius pc, assume first that p, q, r, s are on a semicircle. It follows that one of the angles with vertex c is equal to the sum of the other three. Assume the labeling so that

$$\angle c{:}p, s = \angle c{:}p, q + \angle c{:}q, r + \angle c{:}r, s,$$

where $\angle c{:}x, y$ denotes the angle made by the tangents to the great circles through c, x and c, y, respectively, at their common point c. By the use of simple trigonometric identities, it is seen that

$$\begin{aligned} &[\sin(\tfrac{1}{2}\angle c{:}p, r)\cdot\sin(\tfrac{1}{2}\angle c{:}\ q, r)] + [\sin(\tfrac{1}{2}\angle c{:}p, q)\cdot\sin(\tfrac{1}{2}\angle c{:}r, s)] \\ &\qquad = [\sin(\tfrac{1}{2}\angle c{:}p, r)\cdot\sin(\tfrac{1}{2}\angle c{:}q, s)]. \end{aligned} \tag{2}$$

Multiplication of equation (2) by $\sin^2(pc/\rho)$ and substitution of $\sin(xy/2\rho)$ for $\sin(\frac{1}{2}\angle c{:}x, y)\cdot\sin(pc/\rho)$, where $x, y = p, q, r, s$, lead to the equation

$$[\sin(ps/2\rho)\cdot\sin(qr/2\rho)] + [\sin(pq/2\rho)\cdot\sin(rs/2\rho)] = [\sin(pr/2\rho)\cdot\sin(qs/2\rho)].$$

It follows from the factorization of $\gamma(p, q, r, s)$ as given in the proof of the corollary to Theorem 3.1 that $\gamma(p, q, r, s)=0$.

The only case left is when three of the points lie on a semicircle of $C(c:pc)$ and the other point is on the circle, but not on the same semicircle. This case is treated very much like the above case, which completes the proof.

5. Analogue of the Converse of Ptolemy's Theorem. In order to establish the analogue of the converse of Ptolemy's Theorem, we first establish the following:

THEOREM 5.1. *If p_1, p_2, p_3, p_4, p_5 are five points, then the determinant*

$$\gamma(p_1, p_2, p_3, p_4, p_5) = \left| \sin^2(p_ip_j/2\rho) \right|, \quad \textit{where} \quad i, j = 1, 2, 3, 4, 5 \textit{ vanishes.}$$

Proof. Since $\Delta(p_1, p_2, p_3, p_4, p_5)=0$, it follows that the determinant A, obtained from $\Delta(p_1, p_2, p_3, p_4, p_5)$ by bordering it with a first column and a first row with common element -1, and the rest of the elements of the first column all ones, and the rest of the elements in the first row all zeroes, also vanishes. If the quintuple contains a triple that does not lie on a great circle, then the leading principal minor of order four of A is nonzero, while the determinants obtained from this principal minor by adjoining one row and one column or two rows and two columns vanish. It follows that the rank of A is four. If each triple lies on a great circle, then the rank of A is three. The determinant B, obtained from A by subtracting the first column from the second, third, fourth, fifth, and sixth columns, respectively, and then substituting $-2\sin^2(x/2\rho)$ for $\cos(x/\rho)-1$ also has rank three or four. Hence,

$$\left| \sin^2(p_ip_j/2\rho) \right| = 0, \quad \text{where} \quad i, j = 1, 2, 3, 4, 5.$$

THEOREM 5.2. *If p, q, r, s are four points with $\gamma(p, q, r, s)=0$ and if q, r, s are on a great circle, then p is also on that great circle.*

Proof. Since a point has exactly one diametral point, no generality is lost if it is assumed that q and s are not diametral points. Then q, s determine a unique great circle. If p and any one of q, r, s are diametral points, then p lies on the great circle through q, r, s and the proof is complete. Suppose then, that p is not the diametral point of q, r, or s.

If the first row of the determinant (1) in the proof of Theorem 3.1 is interchanged with the second, third, fourth, and fifth rows; and if $[4, 5]$ denotes the cofactor of the element in the fourth row and fifth column of the determinant so obtained, then it follows that $[4, 5]=0$.

A fairly straightforward expansion of the determinant $[4, 5]$ utilizing the spherical law of cosines as well as the spherical law of sines (see [**1**], pp. 183–185 for this type expansion), yields $\sin(\angle p:q, s)=0$. Consequently, p, q, s lie on a great circle. Since q, r, s lie on a great circle and q, s determine a unique great circle, it follows that p, q, r, s lie on a great circle.

THEOREM 5.3. *If p, q, r, s are four points with $\gamma(p, q, r, s)=0$ and if q, r, s lie on a circle, then p is also on that circle.*

Proof. If q, r, s lie on a great circle and $\gamma(p, q, r, s)=0$, then by Theorem 5.2, p also lies on that great circle. Suppose then that q, r, s lie on a circle C that is not a great circle, $\gamma(p, q, r, s)=0$, and p is not on C. It follows that one of the great circles $C(p, q)$, $C(p, r)$, or $C(p, s)$, say $C(p, q)$ has another point t common to C. Then $\gamma(q, r, s, t, p)=0$ by Theorem 5.1, $\gamma(q, r, s, p)=0$ by hypothesis, and $\gamma(q, r, s, t)=0$ by Theorem 4.1. Hence, the rank of $\gamma(q, r, s, t, p)$ is three. Consequently, $\gamma(r, p, t, q)=0$, and p, t, q are on a great circle. Hence, by Theorem 5.2, q, r, t are on a great circle, contrary to the fact that q, r, t lie on a circle which is not a great circle. Therefore, p lies on the circle of q, r, s.

Since each three points of the spherical plane lie on a circle, the following spherical analogue of Ptolemy's Theorem and its converse has now been established.

THEOREM 5.4. *Four points p_1, p_2, p_3, p_4 of the spherical plane lie on a circle if and only if the determinant*

$$\gamma(p_1, p_2, p_3, p_4) = \left| \sin^2(p_ip_j/2\rho) \right| = 0, \quad \textit{where} \quad i, j = 1, 2, 3, 4.$$

COROLLARY. *If $p_1, p_2, \cdots, p_n$ are n pairwise distinct points of the spherical plane ($n \geqq 4$), then a necessary and sufficient condition that $p_1, p_2, \cdots, p_n$ lie on a circle is that the determinant $\gamma(p_1, p_2, \cdots, p_n)=\left|\sin^2(p_ip_j/2\rho)\right|$, where $i, j=1, 2, \cdots, n$, has rank three.*

Proof. Suppose $p_1, p_2, \cdots, p_n$ are pairwise distinct points of the spherical plane which lie on a circle. Then the leading principal minor of order three of $\gamma(p_1, p_2, \cdots, p_n)$ is nonzero, while the determinants obtained from this principal minor by adjoining one row and one column or two rows and two columns vanish by Theorem 4.1 and Theorem 5.1, respectively. Therefore, the rank of $\gamma(p_1, p_2, \cdots, p_n)$ is three.

Conversely, if the rank of $\gamma(p_1, p_2, \cdots, p_n)$ is three, then every principal minor of $\gamma(p_1, p_2, \cdots, p_n)$ of order four is zero. It follows from Theorem 5.3 that each quadruple of the points, and hence the n points, lie on a circle.

References

1. L. M. Blumenthal, A Modern View of Geometry, Freeman, San Francisco, 1961.
2. ———, Theory and Applications of Distance Geometry, Clarendon Press, Oxford, 1953.
3. ——— and B. E. Gillam, Distribution of points in n-space, this MONTHLY, 50 (1943) 181–185.
4. J. Haantjes, A characteristic local property of geodesics in certain metric spaces, Proc. Akad. Wetensch., Amsterdam, 50 (1947) 496–508.
5. ———, De Stelling von Ptolemeus, Simon Stevin (29), (1952) 25–31.
6. T. Kubota, On the extended Ptolemy's theorem in hyperbolic geometry, Sci. Rep. Tohoku Univ., I (2, 3) (1912) 131–156.

GROUPS WHICH ARE THE UNION OF THREE SUBGROUPS

M. BRUCKHEIMER, The Open University, England

A. C. BRYAN and A. MUIR, The City University, London, England

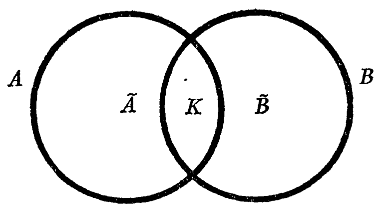
In this paper we consider groups which can be written as the (set-theoretic) union of three subgroups. For completeness, we first dispose of the union of two subgroups; although this result is well known, we use similar methods of proof subsequently. (Throughout this paper we use *union* in the set union sense and not as usually used in group theory.)

LEMMA 1. *Except for trivial cases, a group cannot be the union of two subgroups.*

Proof. Let G be a group, and suppose there exist two subgroups A and B such that

$$A \cup B = G.$$

Clearly $A=G$ or $B=G$ (i.e., $A\subseteq B$ or $B\subseteq A$) represent possible cases: such trivial cases, where one subgroup is a subset of another subgroup, will not be considered in the future. Suppose, therefore, that we have the situation of Fig. 1,



$\tilde{A}=A\setminus B$
$\tilde{B}=B\setminus A$
$K=A\cap B$

FIG. 1

where $\tilde{A}\neq\varnothing$ and $\tilde{B}\neq\varnothing$. Let $\tilde{a}\in\tilde{A}$ and $\tilde{b}\in\tilde{B}$. Then

$$\tilde{a}\tilde{b} \in A \Rightarrow \tilde{b} \in A,$$
$$\tilde{a}\tilde{b} \in B \Rightarrow \tilde{a} \in B,$$

both of which contradict the hypotheses. Hence $\tilde{a}\tilde{b}\notin A\cup B=G$, which contradicts the closure of G.

We now turn to the three group case, and suppose that A, B, and C are three subgroups of G such that

$$A \cup B \cup C = G.$$

If $A\subseteq B$, etc., we are effectively dealing with two subgroups of G, and this is not possible by Lemma 1. Hence we have the configuration of Fig. 2, where $\tilde{A}$, $\tilde{B}$ and $\tilde{C}$ must be nonempty sets.

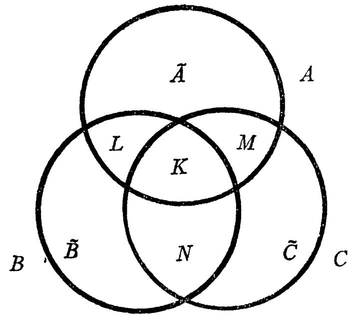


Fig. 2

Lemma 2. $L = M = N = \varnothing$.

Proof. Consider $l \in L$ and $\tilde{c} \in \tilde{C}$. Then

$$\tilde{c}l \in C \Rightarrow l \in C,$$
$$\tilde{c}l \in B \Rightarrow \tilde{c} \in B,$$
$$\tilde{c}l \in A \Rightarrow \tilde{c} \in A,$$

all of which contradict the hypotheses. Since $\tilde{C} \neq \varnothing$, we must have $L = \varnothing$. Similarly $M = N = \varnothing$.

Lemma 3. *Each of $\tilde{A}$, $\tilde{B}$, and $\tilde{C}$ contains its inverses.*

Proof. The inverse $\tilde{a}^{-1}$ of any element $\tilde{a} \in \tilde{A}$ either belongs to $\tilde{A}$ or to K. Suppose $\tilde{a}^{-1} = k \in K$. Then, from $\tilde{a}k = e \in K$, it follows that $\tilde{a} \in K$. Hence $\tilde{a}^{-1} \in \tilde{A}$.

Lemma 4. *If $\tilde{a} \in \tilde{A}$ and $\tilde{b} \in \tilde{B}$, then $\tilde{a}\tilde{b} \in \tilde{C}$.*

The proof follows by contradiction as before.

Lemma 5. *If $\tilde{a}$, $\tilde{a}_1 \in \tilde{A}$, then $\tilde{a}\tilde{a}_1 \in K$.*

Proof. The element $\tilde{a}\tilde{a}_1$ belongs either to $\tilde{A}$ or to K. Suppose that it belongs to $\tilde{A}$, and consider $\tilde{b}\tilde{a}\tilde{a}_1$ for some $\tilde{b} \in \tilde{B}$. Then by Lemma 4, used twice,

$$(\tilde{b}\tilde{a})\tilde{a}_1 \in \tilde{B}.$$

By hypothesis and Lemma 4 again, $\tilde{b}(\tilde{a}\tilde{a}_1) \in \tilde{C}$. Hence $\tilde{a}\tilde{a}_1 \in K$.

Lemma 6. *K is an invariant subgroup of G.*

Proof. Clearly K is a group. Further if $\tilde{a} \in \tilde{A}$ and $k \in K$, then

$$\tilde{a}^{-1}k\tilde{a} \in K.$$

For $k\tilde{a}$ belongs to $\tilde{A}$ or K; since $k\tilde{a} \in K \Rightarrow \tilde{a} \in K$, we must have $k\tilde{a} \in \tilde{A}$. Hence by Lemmas 3 and 5, the result follows.

Clearly the left cosets $\tilde{a}K$ of K are subsets of $\tilde{A}$. In fact

LEMMA 7. $\tilde{a}K = \tilde{A}$.

Proof. Suppose there is an element $\tilde{a}_1 \in \tilde{A}$, but $\tilde{a}_1 \notin \tilde{a}K$. Then by Lemmas 3 and 5, $\tilde{a}^{-1}\tilde{a}_1 = k \in K$. Hence $\tilde{a}_1 = \tilde{a}k$, for some $k \in K$.

It follows from this result and Lemma 4 that the factor group G/K is the Klein Four Group, V.

THEOREM 1. *A group G is the (nontrivial) union of three subgroups if and only if it is homomorphic to the Klein Four Group.*

Proof. One half of this theorem has been proved; it remains to show that if G is homomorphic to V then it is the union of three subgroups. Let K be the kernel of the homomorphism and $\tilde{A}$, $\tilde{B}$, $\tilde{C}$ the cosets of K. Consider the structure A, where

$$A = \tilde{A} \cup K.$$

(i) A is closed, since

$$\left.\begin{aligned} \tilde{A}\tilde{A} &= K \Rightarrow \tilde{a}\tilde{a}_1 \in K \\ \tilde{A}K &= \tilde{A} \Rightarrow \tilde{a}k \in \tilde{A} \end{aligned}\right\} \quad \text{for all } \tilde{a}, \tilde{a}_1 \in \tilde{A} \quad \text{and} \quad k \in K.$$

(ii) $a \in A \Rightarrow a^{-1} \in A$, since each element in V is its own inverse.
This completes the proof.

It follows from Theorem 1, that if G is finite, it must be of order $4m$. Clearly there are many groups of order $4m$ which are not the union of three subgroups. For example the cyclic group C_{4m} of order $4m$ and the alternating group A_n for $n > 4$. Also any locally cyclic group is not the union of three subgroups. This follows from Lemma 5 by considering, say, the subgroup generated by an $\tilde{a}$ and a $\tilde{b}$. This subgroup cannot be cyclic.

On the other hand there are many examples of groups of order $4m$ which are the union of three subgroups. We shall consider a few examples before returning to more general considerations. We shall refer to the *decomposition* of a group into (three) subgroups and call a group which allows such a decomposition a 3-*group*.

Example 1. V itself obviously admits such a decomposition.

$$V \to \{C_2, C_2, C_2\}.$$

Example 2. There are five groups of order 8. Disregarding C_8, we are left with $C_4 \times C_2$, $C_2 \times C_2 \times C_2$, D_4, and Q (the quaternion group). Each admits a decomposition as follows:

$$\begin{aligned} C_4 \times C_2 &\to \{C_4, C_4, V\}, \\ C_2 \times C_2 \times C_2 &\to \{V, V, V\}, \\ D_4 &\to \{C_4, V, V\}, \\ Q &\to \{C_4, C_4, C_4\}. \end{aligned}$$

It will be noticed that there are only four different decompositions possible for a group of order 8, and this displays an example of each. (We consider two decompositions $\{X, Y, Z\}$ and $\{X_1, Y_1, Z_1\}$ as the same if there exist isomorphisms $X \to X_1$, $Y \to Y_1$, $Z \to Z_1$.)

Example 3. The dihedral group D_{2m} of order $4m$ defined by

$$a^{2m} = b^2 = (ab)^2 = e,$$

admits a decomposition for all m. The subgroup K, where

$$K = \{e, a^2, \cdots, a^{2(m-1)}\},$$

is invariant, and the factor group D_{2m}/K is V. The decomposition is

$$D_{2m} \to \{C_{2m}, D_m, D_m\}.$$

Example 4. The dicyclic group of order $4m$, defined by

$$a^{2m} = e, \quad a^m = (ab)^2 = b^2,$$

admits a decomposition for all even m. The subgroup K, where

$$K = \{e, a^2, \cdots, a^{2(m-1)}\},$$

is invariant, and the factor group modulo K is V. The subgroups of the decomposition are C_{2m} and two dicyclic groups, each of order $2m$.

We can immediately dispose of two questions:

(i) Can a 3-group have different decompositions?

(ii) Can two different 3-groups have the same decomposition?

Both questions, as is to be expected, are answered in the affirmative by the following examples.

Example 5. $C_2 \times D_4$ has two distinct decompositions:

$$C_2 \times D_4 \to \{C_4 \times C_2, D_4, D_4\},$$

and

$$C_2 \times D_4 \to \{C_4 \times C_2, C_2 \times C_2 \times C_2, C_2 \times C_2 \times C_2\}.$$

Example 6. The group of order 16 defined by the generating relations

$$a^2 = b^2 = c^2 = e, abc = bca = cab,$$

has a decomposition $\{C_4 \times C_2, D_4, D_4\}$; from Example 5, this is also a decomposition of $C_2 \times D_4$.

If a 3-group is non-Abelian, then the subgroups of a decomposition can be (locally) Abelian ($\mathfrak{A}$) or non-Abelian ($\mathfrak{N}$). This gives four possible types of decompositions, namely, $\{\mathfrak{A}, \mathfrak{A}, \mathfrak{A}\}$, $\{\mathfrak{A}, \mathfrak{A}, \mathfrak{N}\}$, $\{\mathfrak{A}, \mathfrak{N}, \mathfrak{N}\}$, and $\{\mathfrak{N}, \mathfrak{N}, \mathfrak{N}\}$. The decomposition $\{\mathfrak{A}, \mathfrak{A}, \mathfrak{N}\}$, however, cannot exist as is shown by

LEMMA 8. *If $G \to \{A, B, C\}$ and A and B are Abelian, then C is also Abelian.*

Proof. Since $K \subset A$, K is Abelian. If $\tilde{c}, \tilde{c}_1 \in \tilde{C}$ then we can always write $\tilde{c} = \tilde{a}\tilde{b}$

and $\tilde{c}_1 = \tilde{a}\tilde{b}_1 = \tilde{a}\tilde{b}k$, for some $\tilde{a} \in \tilde{A}$; $\tilde{b}$, $\tilde{b}_1 \in \tilde{B}$ and $k \in K$. It follows that $\tilde{C}$ is commutative. Similarly the elements of $\tilde{C}$ and K commute.

The remaining three types of decompositions all exist as the following example shows.

Example 7.

$$Q \to \{C_4,\ C_4,\ C_4\} \quad \text{is of form } G \to \{\mathfrak{A},\ \mathfrak{A},\ \mathfrak{A}\},$$
$$D_6 \to \{C_6,\ S_3,\ S_3\} \quad \text{is of form } G \to \{\mathfrak{A},\ \mathfrak{N},\ \mathfrak{N}\},$$
$$S_3 \times V \to \{D_6,\ D_6,\ D_6\} \quad \text{is of form } G \to \{\mathfrak{N},\ \mathfrak{N},\ \mathfrak{N}\}.$$

We summarize these results in the following theorem:

THEOREM 2. *Each decomposition of a* 3-*group is one of the forms* $\{\mathfrak{A},\ \mathfrak{A},\ \mathfrak{A}\}$, $\{\mathfrak{A},\ \mathfrak{N},\ \mathfrak{N}\}$ *or* $\{\mathfrak{N},\ \mathfrak{N},\ \mathfrak{N}\}$.

If a 3-group G is Abelian then its center is G. For non-Abelian 3-groups we have the following two results:

THEOREM 3. *A non-Abelian* 3-*group* G *has an Abelian decomposition, i.e.,* $G \to \{\mathfrak{A},\ \mathfrak{A},\ \mathfrak{A}\}$, *if and only if the center of* G *is* K.

Proof. Let $G \to \{A,\ B,\ C\}$ and let Z be the centre of G.

Firstly, assume that A, B, and C are Abelian. Then $Z \supseteq K$. Suppose $\tilde{a}_z \in \tilde{A} \cap Z$. For each $\tilde{c} \in \tilde{C}$ there exists a $\tilde{b} \in \tilde{B}$ such that $\tilde{c} = \tilde{a}_z\tilde{b}$; whence $\tilde{c}b = b\tilde{c}$, for all $b \in B$. Since $Z \supseteq K$, it follows that the elements of B and C commute.

Also, for each $\tilde{a} \in \tilde{A}$ there exists some element $k \in K$ such that $\tilde{a} = \tilde{a}_z k$, so that $\tilde{a}b = b\tilde{a}$, for all $b \in B$. Thus the elements of A and B commute. Similarly, we can show that the elements of A and C commute.

It follows that G is Abelian, which is a contradiction. Hence $Z = K$.

Conversely, suppose $Z = K$. For any fixed element $\tilde{a}_1 \in \tilde{A}$, there exist elements $k \in K$ such that $\tilde{a} = \tilde{a}_1 k$, for all $\tilde{a} \in \tilde{A}$. It follows that A is Abelian. Similarly B and C are Abelian.

THEOREM 4. *If* G *admits a decomposition* $\{A,\ B,\ C\}$ *of type* $\{\mathfrak{A},\ \mathfrak{N},\ \mathfrak{N}\}$, *then the center* Z *of* G *is contained in* A.

Proof. Suppose $\tilde{b}_z \in \tilde{B} \cap Z$. For every $\tilde{b} \in \tilde{B}$ there exists *a* $k \in K$ such that $\tilde{b} = \tilde{b}_z\, k$.

It follows that the elements of $\tilde{B}$ and K commute, and that $\tilde{B}$ itself is commutative. Therefore B is Abelian, which is a contradiction. Similarly $\tilde{C} \cap Z = \emptyset$.

Example 8. For $D_6 \to \{C_6,\ S_3,\ S_3\}$, the center Z is such that $Z \cap \tilde{A} \neq \emptyset$.

From Theorems 3 and 4 we have that if G has a decomposition $\{A,\ B,\ C\}$ of type $\{\mathfrak{A},\ \mathfrak{N},\ \mathfrak{N}\}$, then the center Z is either properly contained in K or, at most, it also contains elements of $\tilde{A}$. Are there other possibilities when the decomposition is of type $\{\mathfrak{N},\ \mathfrak{N},\ \mathfrak{N}\}$? It is clear that Z cannot contain elements of $\tilde{A}$ and $\tilde{B}$ and not $\tilde{C}$, since $Z \cap A$ and $Z \cap B$ are groups and a group cannot be expressed as the nontrivial union of two groups (Lemma 1). But there is the possibility that Z contains elements of $\tilde{A}$, $\tilde{B}$ and $\tilde{C}$; in this case Z is itself neces-

sarily a 3-group with decomposition $Z \rightarrow \{Z \cap A, Z \cap B, Z \cap C\}$. ($Z$ may, of course, be a 3-group in other cases.) The existence of such a decomposition requires that K be non-Abelian. For, if K were Abelian we could prove that, for instance, A was Abelian. The method of proof is similar to that used in Theorem 4.

Example 9. For $S_3 \times V \rightarrow \{D_6, D_6, D_6\}$, none of $Z \cap \tilde{A}$, $Z \cap \tilde{B}$, $Z \cap \tilde{C}$ is empty.

Let $\bar{I}(A)$ be the set of inner automorphisms of G defined by elements of A. Then $\bar{I}(A)$ is a subgroup of $I(G)$, the group of inner automorphisms of G; we have, trivially, $I(G) = \bar{I}(A) \cup \bar{I}(B) \cup \bar{I}(C)$. Therefore, we have

THEOREM 5. *The group of inner automorphisms of a* 3-*group G is either itself a* 3-*group or degenerate, in the sense that it is one of* $\bar{I}(A)$, $\bar{I}(B)$, *or* $\bar{I}(C)$.

We note that degeneracy (as defined in Theorem 5) does not necessarily preclude $I(G)$ from being a 3-group, as is shown by the next example. However, as we are concerned with the structure of $I(G)$ relative to $G \rightarrow \{A, B, C\}$, this definition is appropriate.

Example 10.

$Q \rightarrow \{C_4, C_4, C_4\}$;	$I(G) = V$	(nondegenerate)
$D_6 \rightarrow \{C_6, S_3, S_3\}$;	$I(G) = \bar{I}(S_3) = S_3$	(degenerate)
$C_2 \times D_4 \rightarrow \{C_4 \times C_2, D_4, D_4\}$;	$I(G) = \bar{I}(D_4) = V$	(degenerate)
$C_2 \times D_4 \rightarrow \{C_4 \times C_2, C_2 \times C_2 \times C_2, C_2 \times C_2 \times C_2\}$;	$I(G) = V$	(nondegenerate).

For an Abelian decomposition of a non-Abelian 3-group G, it follows from Theorems 1 and 3 that $I(G) = G/Z = V$. Hence, we have

THEOREM 6. *A non-Abelian* 3-*group has an Abelian decomposition if and only if the group of inner automorphisms is the Klein Four Group.*

Finally, we consider the relation between the 3-group structure and degeneracy of $I(G)$. This is given by

THEOREM 7. *The group of inner automorphisms of a* 3-*group is degenerate if and only if the center contains elements other than elements from K.*

Proof. Suppose $\tilde{a}_z \in \tilde{A} \cap Z$. Since for every $\tilde{b} \in \tilde{B}$ there are elements $\tilde{c} \in \tilde{C}$ such that $\tilde{b} = \tilde{c}\tilde{a}_z$. If we denote by $i(g)$ the inner automorphism of G defined by g, then we have that $i(\tilde{b}) = i(\tilde{c})$. Therefore $\bar{I}(\tilde{B}) = \bar{I}(\tilde{C})$, and $\bar{I}(B) = \bar{I}(\tilde{B}) \cup \bar{I}(K) = \bar{I}(C)$. Hence $I(G)$ is degenerate.

In fact, since every element $\tilde{a} \in \tilde{A}$ is of the form $\tilde{a} = \tilde{a}_z k$, for some $k \in K$, we have $i(\tilde{a}) = i(k)$. Therefore $\bar{I}(A) = \bar{I}(K)$ and $I(G) = \bar{I}(B)$.

Conversely, $\bar{I}(B) \subseteq \bar{I}(C)$ implies either

$$\text{(i) } i(\tilde{b}) = i(k) \Rightarrow \tilde{b} = kZ \Rightarrow Z \cap \tilde{B} \neq \emptyset,$$

or

$$\text{(ii) } i(\tilde{b}) = i(\tilde{c}) \Rightarrow \tilde{b} = \tilde{c}Z \Rightarrow Z \cap \tilde{A} \neq \emptyset.$$

CORRECTION TO "FUNCTIONAL ANALYSIS PROOFS OF SOME THEOREMS IN FUNCTION THEORY"

L. A. RUBEL AND B. A. TAYLOR (vol. 76 (1969) 483–489)

There is a gap, pointed out to the authors by Benjamin Lepson, in the argument supporting the Remark on page 486, since h may have unwanted zeros. To fill this gap, simply take $f=(g/h)k$, where g and h are as before, but where k has a high order zero at each unwanted zero of h, and where k takes the value 1, to a high order, at each wanted zero of h.

MATHEMATICAL NOTES

EDITED BY DAVID DRASIN

Manuscripts for this Department should be sent to David Drasin, Division of Mathematical Sciences, Purdue University, Lafayette, IN 47907.

OPTIMALLY SEPARATED CONTRACTIONS

D. J. NEWMAN, Yeshiva University, and L. RAYMON, Temple University

We introduce the class of (normalized) contractions, K, as the set of functions, $f(x)$, defined on $[0, 1]$ and satisfying

$$f(0) = 0, \qquad |f(x) - f(y)| \leqq |x - y| \quad \text{for all } x, y \text{ in } [0, 1].$$

Under the usual sup-norm metric,

$$\|f(x)\| = \max_{0 \leqq x \leqq 1} |f(x)|,$$

the class K forms a compact set of functions and one which plays a central role in approximation theory. Our purpose in this note is to determine how far apart n elements of K can be from one another.

THEOREM I. *Given that* $f_1, f_2, \cdots, f_n \in K$ *we can choose* $i \neq j$ *such that*

$$\|f_i - f_j\| \leqq 2/\langle \log_2 n \rangle.$$

II. *There exist* n *functions* $f_1, f_2, \cdots, f_n \in K$ *such that for all* i, j $(i \neq j)$ *we have*

$$\|f_i - f_j\| \geqq 2/\langle \log_2 n \rangle.$$

(Here $\langle x \rangle$ denotes the smallest integer which isn't smaller than x; e.g., $\langle 5 \rangle = 5$, $\langle \pi \rangle = 4$, $\langle -\sqrt{2} \rangle = -1$.)

Proof. Clearly we need to prove I only for $n = 2^k + 1$ and II only for $n = 2^k$. The basic consideration in either case is the collection of functions $\phi_1, \phi_2, \cdots, \phi_{2^k}$, namely those continuous functions which vanish on the negative axis and whose slope is constantly $+1$ or -1 in each of the intervals $((\nu-1)/k, \nu/k)$, where ν is any positive integer not exceeding k.

Proof of II. For our 2^k functions f_i we choose the functions ϕ_i themselves. Clearly these belong to K. If $i \neq j$, moreover, then there is a first interval $((\nu-1)/k, \nu/k)$ wherein the slopes of ϕ_i and ϕ_j are different (otherwise ϕ_i would identically equal ϕ_j, contradicting the fact that $i \neq j$). We may suppose that ϕ_i has slope $+1$, ϕ_j slope -1 in this interval, then

$$\begin{aligned}\phi_i((\nu - 1)/k) &= \phi_j((\nu - 1)/k),\\ \phi_i(\nu/k) &= \phi_i((\nu - 1)/k) + 1/k,\\ \phi_j(\nu/k) &= \phi_j((\nu - 1)/k) - 1/k.\end{aligned}$$

Hence $\phi_i(\nu/k) - \phi_j(\nu/k) = 2/k$, and we conclude that $\|\phi_i - \phi_j\| \geqq 2/k$ as required.

For I we need the following fact:

LEMMA. *If $f \in K$ then*

$$\| f(x) - k(k+1)^{-1}\phi_m(((k+1)x - 1)/k)\| \leqq 1/(k+1)$$

for at least one of our functions ϕ_m.

Proof. $f(x)$ is given. Define a sequence of ± 1's inductively by setting, for $\nu = 1, 2, \cdots, k$,

$$\epsilon_\nu = sg[(k+1)f((\nu+1)/(k+1)) - (\epsilon_1 + \epsilon_2 + \cdots + \epsilon_{\nu-1})]$$

(the empty sum being interpreted as 0) where we define $sgx = +1$ for $x \geqq 0$ and -1 for $x < 0$.

Choose ϕ_m so as to have the slopes ϵ_ν throughout $[(\nu-1)/k, \nu/k]$. We claim that this choice satisfies our requirements. We simply proceed by induction to prove that, for $\nu = 0, 1, 2, \cdots, k$,

$$| f(x) - k(k+1)^{-1}\phi_m(((k+1)x - 1)/k) | \leqq 1/(k+1)$$

throughout $[\nu/(k+1), (\nu+1)/(k+1)]$. This is clearly true for $\nu = 0$ so we assume it to hold below $\nu \leqq k$ and proceed to the case of ν itself.

By definition of ϕ_m, we have, throughout $[\nu/(k+1), (\nu+1)/(k+1)]$,

$$k(k+1)^{-1}\phi_m(((k+1)x - 1)/k) = (\epsilon_1 + \epsilon_2 + \cdots + \epsilon_{\nu-1})/(k+1) + (x - \nu/(k+1))\epsilon_\nu.$$

Hence the required inequality can be written as $-1 \leqq g(x) \leqq 1$ throughout $[\nu/(k+1), (\nu+1)/(k+1)]$, where

$$g(x) = [(k+1)f(x) - (\epsilon_1 + \epsilon_2 + \cdots + \epsilon_{\nu-1})]/\epsilon_\nu - ((k+1)x - \nu).$$

Now note that the function $g(x)$ is *nonincreasing*. It is therefore sufficient to show that

$$g(\nu/(k+1)) \leqq 1, \qquad g((\nu+1)/(k+1)) \geqq -1.$$

The first inequality follows by the inductive hypothesis since

$$g(\nu/(k+1)) = [(k+1)f(\nu/(k+1)) - (\epsilon_1 + \cdots + \epsilon_{\nu-1})]/\epsilon_\nu.$$

As for the second inequality we have

$$g((\nu+1)/(k+1)) = \frac{(k+1)f((\nu+1)/(k+1)) - (\epsilon_1 + \cdots + \epsilon_{\nu-1}) - 1}{\epsilon_\nu}$$

and this is $\geqq -1$, since, by the definition of ϵ_ν,

$$[(k+1)f((\nu+1)/(k+1)) - (\epsilon_1 + \cdots + \epsilon_{\nu-1})]/\epsilon_\nu \geqq 0.$$

We can now give the

Proof of I. Let $f_1, f_2, \cdots, f_{2^k+1}$ be given and for each of these f's determine the corresponding ϕ_m given by the lemma. For at least two of these, say f_i and f_j, we obtain the same ϕ_m. Thus

$$\|f_i(x) - k(k+1)^{-1}\phi_m([(k+1)x-1]/k)\| \leqq 1/(k+1) \text{ and}$$
$$\|f_j(x) - k(k+1)^{-1}\phi_m([(k+1)x-1]/k)\| \leqq 1/(k+1) \text{ so that}$$
$$\|f_i(x) - f_j(x)\| \leqq 2/(k+1).$$

Partially supported by NSF GP 7504.

CONCERNING THE ITERATED ϕ FUNCTION

P. A. CATLIN, Carnegie-Mellon University

Define a set of functions ϕ^n mapping integers greater than 1 into the natural numbers, such that $\phi^0(a)=a$ and $\phi^{n+1}(a)=\phi(\phi^n(a))$, where ϕ is Euler's function. Furthermore, define the class $C(a)$ of a to be the integer n such that $\phi^n(a)=2$, and let M be the set of natural numbers with the least value of any member in their respective classes. Harold Shapiro made the conjecture in the last paragraph of his article [2] that each element of M was prime, and W. H. Mills [1] exhibited several composite members of M. The purpose of this note is to extend their results concerning the factorization of elements of M.

H. Shapiro gave a theorem (Theorem 15, [2]) which stated essentially that if S is the set of numbers in class n that are less than 2^{n+1}, then the factors of an element of S are in S. This is also true for odd elements of M.

THEOREM 1. *If m is an odd element of M, the factors of m are in M.*

Proof. It is sufficient to show that an arbitrary factor is in M. If $m \in M$ is prime, then the theorem is obvious. Otherwise, if $m \in M$ is odd, then so are its factors. Let $m=ab$. By Theorem 1 of [2], $C(m)=C(a)+C(b)$.

If $b \notin M$ then there must be a lower number $c \in M$ in the same class as b. However, by Theorem 1 of [2], we see that $C(m)=C(a)+C(b)=C(ac)$ where $ac<m$, contradicting the definition of M. Thus, $b \in M$, proving the theorem.

As a consequence of Theorem 7 of [2], we can make a similar statement for even members of M: in this case, the only prime factor of an element of M is 2, which is in M. However, 2 is the only known even member of M.

The following theorem shows that the existence of finitely many primes in M is equivalent to an apparently weaker condition:

THEOREM 2. *There are finitely many primes in M if and only if there are finitely many odd numbers in M.*

Proof. If there are infinitely many primes in M, then obviously there are infinitely many odd numbers. Conversely, if an odd prime $p \in M$ then $p^\alpha \in M$ for at most a finite number of α. This follows since $p^\alpha \in M$ implies

$$2^{\alpha C(p)} = 2^{C(p^\alpha)} < p^\alpha < 2^{\alpha C(p)+1}$$

or

$$2^{C(p)} < p < 2^{C(p)+1/\alpha}$$

which is false for α sufficiently large. Thus, if M contains only finitely many odd primes then it contains only finitely many odd prime powers; and by Theorem 1, only finitely many odd integers.

As a corollary, Theorem 2 implies that S contains infinitely many primes if and only if it contains infinitely many odd numbers. This follows because for any $m \in M$ there are finitely many $s \in S$ such that $m \leqq s < 2^{C(s)+1}$, so if M contains finitely many primes, so does S; the converse follows from $M \subset S$.

The questions of whether or not M and S contain infinitely many odd integers remain open. In all likelihood the smallest integer of each class is odd.

References

1. W. H. Mills, Iteration of the ϕ function, this MONTHLY, 50 (1943) 547–549.
2. Harold Shapiro, An arithmetic function arising from the ϕ function, this MONTHLY, 50 (1943) 18–30.

A NOTE ON COMMUTING AUTOMORPHISMS OF RINGS

JIANG LUH, North Carolina State University at Raleigh

Let R be a ring. An automorphism T of R is called a *commuting automorphism* of R if $x^T x = x x^T$ for each x in R. If, in addition, T is not the identity automorphism of R, then T is called a *nontrivial commuting automorphism*. In [1], Divinsky has proved that if a simple Artinian ring possesses a nontrivial commuting automorphism, then it is a field.

The purpose of this note is to generalize the Divinsky result for arbitrary prime rings. The proof here is simpler and shorter than that given by Divinsky for the simple Artinian case.

THEOREM. *Let R be a prime ring. If R possesses a nontrivial commuting automorphism, then R is a commutative integral domain.*

As an immediate consequence of the theorem, we have

COROLLARY. *Let R be a primitive ring. If R possesses a nontrivial automorphism, then R is a field.*

To prove the theorem, we shall require the following lemma due to Divinsky [1]. For easy reference, we exhibit his proof here.

LEMMA 1. *If T is a commuting automorphism of a ring R, then for each $x, y \in R$, $(x-x^T)[x, y]=0$, where $[x, y]=xy-yx$.*

Proof. Polarizing $[x, x^T]=0$ gives $[x, y^T]=[x^T, y]$ and hence $[x, (xy)^T] = [x^T, xy]$. But $[x, (xy)^T]=x^T[x, y^T]$ and $[x^T, xy]=x[x^T, y]=x[x, y^T]$. Thus, $(x-x^T)[x, y^T]=0$. Replacing y^T by y since T is an automorphism, we establish the lemma.

COROLLARY 1. *If T is a commuting automorphism of a ring R, then for every $x, y \in R$, $(x-x^T)R[x, y]=0$.*

Proof. We note $z[x, y]=[x, zy]-[x, z]y$ for $z \in R$ and use Lemma 1.

COROLLARY 2. *Let R be a prime ring and T a commuting automorphism of R. If $x \in R$, $x \neq x^T$, then $x \in C$, the center of R.*

Proof. By Corollary 1, $(x-x^T)R[x, y]=0$ for all $y \in R$. Since $x-x^T \neq 0$, the primeness of R implies $[x, y]=0$ for all $y \in R$.

Proof of the theorem. Since T is nontrivial, there exists $x \in R$ such that $x \neq x^T$ and by Corollary 2, $x \in C$. Suppose there exists $y \in R$, $y \notin C$; then $x+y \notin C$, and by Corollary 2, $y^T=y$ and $(x+y)^T=x+y$; so $x^T=x$, a contradiction. This completes the proof.

This work was done while the author was at Wright-Patterson Air Force Base under the contract NO. F 33615-67-C-1758 of Ohio State University.

Reference

1. N. Divinsky, On commuting automorphism of rings, Trans. Roy. Soc. Canada, Sect. III, 49 (1955) 19–22.

EIGENVALUES OF A MATRIX OF RANK k

LOUIS BRAND, University of Houston

In an $n \times n$ matrix A of rank k *all* determinants of order $>k$ are zero; hence its characteristic equation is

$$\lambda^n - a_1 \lambda^{n-1} + a_2 \lambda^{n-2} - \cdots + (-1)^k a_k \lambda^{n-k} = 0,$$

where a_r is the sum of all r-rowed *principal* minors of A. Thus A has the eigenvalue 0 at least of multiplicity $n-k$; and its k remaining eigenvalues are roots of

$$\lambda^k - a_1 \lambda^{k-1} + a_2 \lambda^{k-2} - \cdots + (-1)^k a_k = 0,$$

some of which may also be 0.

Since a_1 is the trace of A, the result in [1] follows from this result for $k=1$.

Reference

1. Larry Cummings, Characteristic roots of rank one matrices, this MONTHLY, 75 (1968) 1105–1106.

RESEARCH PROBLEMS

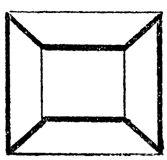
EDITED BY VICTOR KLEE

In this Department the Monthly presents easily stated research problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Material should be sent to Victor Klee, Department of Mathematics, University of Washington, Seattle, WA 98105.

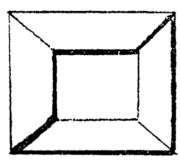
WHAT IS THE MAXIMUM LENGTH OF A d-DIMENSIONAL SNAKE?

VICTOR KLEE, University of Washington

Let $I(d)$ denote the graph formed by the vertices and edges of a d-dimensional cube. The vertices are the 2^d d-tuples of 0's and 1's, and two vertices are joined by an edge if and only if they differ in exactly one coordinate. As the term is used here, a *d-dimensional snake* is a simple circuit C formed from vertices and edges of $I(d)$ such that any edge of $I(d)$ joining two vertices of C is in fact an edge of C. The figures below show two circuits in $I(3)$, the first being of length 8 and the second of length 6. The second is a snake but the first is not, as the defining condition for snakes is violated by each of the four edges omitted from the first circuit. In fact, no 3-dimensional snake has more than 6 vertices.



nonsnake



snake

The title problem asks for an evaluation (or, lacking that, for a good estimate of asymptotic behavior as $d \to \infty$) of $S(d)$, the maximum length of a d-dimensional snake. $S(d)$ is known exactly for only a few values of d; specifically, $S(2)=4$, $S(3)=6$, $S(4)=8$, $S(5)=14$, and $S(6)=26$. Beyond that, it is known that

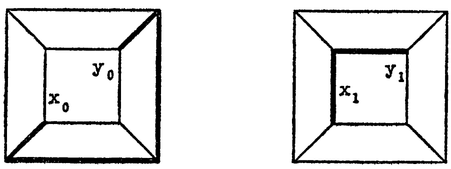
$$(*) \qquad \frac{7}{4(d-1)} \leqq \frac{S(d)}{2^d} \leqq \frac{1}{2} - \frac{1-12/2^d}{7d(d-1)^2+2} \qquad \text{for all } d \geqq 6.$$

The lower bound is due to Danzer and Klee [4], the upper bound to Douglas [7]. The value of $S(5)$ was found by Even [8] and follows from a more general result of Singleton [15] and Douglas [6] stated below. The value of $S(6)$ was found by Davies [5] as the result of a computer search.

Snakes were first studied by Kautz [10], who called them *unit-distance error-checking codes* because they can be used to add an error-checking feature to certain analog-to-digital conversion systems. Žuravlev [17] noted a relationship between the values of $S(\cdot)$ and the efficiency of certain algorithms for simplifying disjunctive normal forms in Boolean algebra. Black [2] suggested a use for

snakes in the design of electronic combination locks. Estimates for $S(d)$ that were superseded by (*) were given by Abbott [1], Danzer and Klee [4], Glagolev [9], Kautz [10], Larman [13], Ramanujacharyulu and Menon [14], Singleton [15], and Vasil'ev [16]. An expository account was given by Klee [12].

In addition to the work on snakes, some attention has been devoted to more restricted classes of circuits that have better error-checking properties. A *d-dimensional circuit code of spread s* (also called *SIB_s code* or *circuit code of minimum distance s*) is a simple circuit C in $I(d)$ such that any two vertices of C differing in exactly r coordinates, with $r<s$, can be joined by a path formed from r edges of C. Every circuit is of spread 1, and the circuits of spread 2 are precisely the snakes. The snake in the figure above is also of spread 3, but the figure below depicts a 4-dimensional snake of length 8 that is not of spread 3, as is seen by considering the vertices x_0 and y_0. In fact, no 4-dimensional circuit of spread 3 has more than 6 vertices.



[$I(4)$ is represented by two copies of $I(3)$, and the eight edges of $I(4)$ not shown join a vertex in one copy to the corresponding vertex in the other copy. The snake in question is formed from the edges x_0x_1 and y_0y_1 in addition to the six heavy edges.]

Let $C(d, s)$ denote the maximum length of circuits in $I(d)$ of spread s. Then $C(d, 1)=2^d$, $C(d, 2)=S(d)$, and for $s>2$ the existing knowledge about $C(d, s)$ is even more meager than for $s=2$. The following values are among those established by Singleton [15] and Douglas [6]: $C(d, s)=2d$ for $d<[3s/2]+2$; $C([3s/2]+2, s)=4s+6$ for even s; $C([3s/2]+2, s)=4s+4$ for odd s; $C([3s/2]+3, s)=4s+8$ for odd $s\geqq 9$. For general d and s, lower bounds on $C(d, s)$ have been given by Singleton [15] and Klee [11], upper bounds by Chien, Freiman, and Tang [3] and Douglas [7].

Preparation of this paper was supported in part by the Office of Naval Research and the Boeing Scientific Research Laboratories.

References

1. H. L. Abbott, Some problems in combinatorial analysis, Ph.D. thesis, Univ. of Alberta, Edmonton, Canada, 1965.

2. W. L. Black, Electronic combination lock, Quart. Progress Report of the Research Laboratory of Electronics, No. 73, Mass. Inst. Tech., Cambridge, Mass., April 1964, p. 232.

3. R. T. Chien, C. V. Freiman, and D. T. Tang, Error correction and circuits on the n-cube, Proc. Second Annual Allerton Conf. on Circuit and System Theory, 1964, Univ. of Illinois Allerton House, Monticello, Ill., pp. 899–912.

4. L. Danzer and V. Klee, Lengths of snakes in boxes, J. Combinatorial Theory, 2 (1967) 258–265.

5. D. W. Davies, Longest "separated" paths and loops in an N cube, IEEE Trans. Electronic Computers, EC14 (1965) 261.

6. R. J. Douglas, Some results on the maximum length of circuits of spread k in the d-cube, J. Combinatorial Theory, 6 (1969) 323–339.

7. ———, Upper bounds on the length of circuits of even spread in the d-cube, J. Combinatorial Theory, 7 (1969) 206–214.

8. S. Even, Snake-in-the-box codes, IEEE Trans. Electronic Computers, EC12 (1963) 18.

9. V. V. Glagolev, An upper estimate of the length of a cycle in the n-dimensional unit cube (Russian), Diskret. Analiz., No. 6 (1966) 3–7.

10. W. H. Kautz, Unit-distance error-checking codes, IRE Trans. Electronic Computers, 7 (1958) 179–180.

11. V. Klee, A method for constructing circuit codes, J. Assoc. Computing Machinery, 14 (1967) 520–528.

12. ———, Long paths and circuits on polytopes. In B. Grünbaum's Convex Polytopes, Wiley, New York, 1967, pp. 356–389.

13. D. G. Larman, Circuit codes, unpublished manuscript, 1968.

14. C. Ramanujacharyulu and V. V. Menon, A note on the snake-in-the-box problem, Publ. Inst. Statist. Univ. Paris, 13 (1964) 131–135.

15. R. C. Singleton, Generalized snake-in-the-box codes, IEEE Trans. Electronic Computers, EC15 (1966) 596–602.

16. Yu. L. Vasil'ev, On the length of a cycle in an n-dimensional unit cube, Soviet Math. Dokl., 4 (1963) 160–163. Translated from Dokl. Akad. Nauk SSSR, 148 (1963) 753–756.

17. Yu. L. Žuravlev, Set-theoretic methods in Boolean algebra (Russian), Probl. Kibernet. No. 8 (1962) 5–44.

CLASSROOM NOTES

EDITED BY DAVID DRASIN

Manuscripts for this Department should be sent to David Drasin, Division of Mathematical Sciences, Purdue University, Lafayette, IN 47907.

ON THE WRONSKIAN TEST FOR INDEPENDENCE

R. A. HORN, The Johns Hopkins University

Let $n \geqq 1$ be an integer and let $\{u_k(x)\}_{k=1}^n$ be n functions each of which is $n-1$ times differentiable on some interval $I=(a, b) \subset \mathbb{R}$. In courses on elementary differential equations the *Wronskian matrix*

$$K(u_1, \cdots, u_n; x) \equiv \begin{bmatrix} u_1(x) & \cdots & u_n(x) \\ u_1^{(1)}(x) & \cdots & u_n^{(1)}(x) \\ \vdots & & \vdots \\ u_1^{(n-1)}(x) & \cdots & u_n^{(n-1)}(x) \end{bmatrix}, \qquad x \in I$$

and the *Wronskian* $W(u_1, \cdots, u_n; x) \equiv \det K(u_1, \cdots, u_n; x)$ are introduced in order to test whether a set of solutions of a normalized homogeneous linear differential equation of order n is a linearly independent set on I. Using the differential equation, one shows that the Wronskian of a set of solutions either

vanishes identically or never vanishes on I. Thus, a set of functions $\{u_k(x)\}_{k=1}^n$ is linearly dependent on I if and only if $W(u_1, \cdots, u_n; x)=0$ for all $x \in I$, *provided that there is some one* nth *order normalized homogeneous linear differential equation on* I *of which each* $u_k(x)$ *is a solution*, $1 \leqq k \leqq n$.

Quite naturally, the alert student will want to know if the latter condition can be dropped so that the identical vanishing of the Wronskian can be used as a test for linear dependence of a general set of (sufficiently smooth) functions. Some textbooks anticipate this question by providing an example to show that it cannot be dropped; a typical example is $u_1(x)=x^2$, $u_2(x)=x|x|$ on $I=(-1, 1)$. The student, however, frequently feels that the example is "unnatural" and that for the usual smooth functions one meets in calculus books the identical vanishing of the Wronskian on I should be a sure test for dependence on I. This feeling is justified by the following easy result.

THEOREM. *Let* $I=(a, b) \subset \mathcal{R}$ *and let* $\{u_k(x)\}_{k=1}^n$ *be a set of* n *analytic functions on* I, $n \geqq 1$. *These functions are linearly dependent on* I *if and only if* $W(u_1, \cdots, u_n; x)=0$ *for all* $x \in I$.

Proof. The "only if" statement is well known and does not require analyticity. Analyticity is essential for the converse, however, as the above example shows. Since the theorem is trivial for $n=1$, we may assume $n \geqq 2$ and by selection of a suitable subset of functions and reordering we may also assume that $W(u_1, \cdots, u_{n-1}; x)$ is not identically zero on I. Thus, there is some nonempty interval $I'=(c, d) \subset I$ such that $W(u_1, \cdots, u_{n-1}; x) \neq 0$ for all $x \in I'$, and hence

$$\frac{W(u_1, \cdots, u_{n-1}, u; x)}{W(u_1, \cdots, u_{n-1}; x)} \equiv u^{(n-1)}(x) + P_1(x)u^{(n-2)}(x) + \cdots + P_{n-1}(x)u(x) = 0$$

is a normalized homogeneous linear differential equation of order $n-1$ with continuous coefficients on I'. Since there can be no more than $n-1$ linearly independent solutions on I' and since each of the n functions $u_1(x), \cdots, u_n(x)$ is a solution, we conclude that these functions must be dependent on I'. Thus, there exist constants $c_1, \cdots, c_n$ such that $f(x) \equiv c_1u_1(x) + \cdots + c_nu_n(x)=0$ for all $x \in I'$. But $f(x)$ is analytic on I and vanishes identically on a nonempty open set $I' \subset I$, so $f(x)=0$ for all $x \in I$. Thus, the set $\{u_k(x)\}_{k=1}^n$ is dependent on I.

If $W(u_1, \cdots, u_n; x) \equiv 0$ on I and if $W(u_1, \cdots, u_{n-1}; x) \neq 0$ for all $x \in I$, we see from the proof that analyticity is not required in order to conclude that the functions are dependent on I. This generalizes the familiar fact that if $W(u_1, u_2; x) \equiv 0$ on I and if $u_1(x) \neq 0$ for all $x \in I$, then u_1 and u_2 are dependent.

For the validity of the theorem it would, of course, be sufficient to assume only that $W(x)=0$ for all x in some nonempty open subset of I, or that $W(x)=0$ for all x in some infinite subset of I with an accumulation point interior to I, or to make any other assumption about $W(x)$ which forces it to vanish everywhere on I. The same proof also shows that the Wronskian of a finite set of functions analytic in some region of the complex plane will vanish identically if and only if the functions are dependent in the region.

A STRONGER VERSION OF A METRIZATION THEOREM OF K. MORITA

J. H. WESTON, J. R. SHILLETO, AND S. A. RANKIN, University of Saskatchewan, Regina

In 1955, the following metrization theorem was proven by K. Morita [2].

THEOREM (Morita). *A T_1 space is metrizable if and only if there exists a sequence $\{\mathfrak{F}_n \mid n$ a positive integer$\}$ of closed, locally finite coverings of X such that for any open set U and $x \in U$ there is a positive integer n with $\bigcup\{F \in \mathfrak{F}_n \mid x \in F\} \subseteq U$.*

The proof which appears in [1] suggests that the locally finiteness condition may be weakened. In fact, we make the following observation.

DEFINITION. *A collection $\mathfrak{A}$ of subsets of a topological space is called closure preserving if and only if for each subcollection $\mathfrak{B}$ of $\mathfrak{A}$, $(\bigcup\{B \mid B \in \mathfrak{B}\})^- = \bigcup\{B^- \mid B \in \mathfrak{B}\}$.*

DEFINITION. *For two collections of sets $\mathfrak{A}$ and $\mathfrak{B}$, let*

$$\mathfrak{A} \wedge \mathfrak{B} = \{A \cap B \mid A \in \mathfrak{A}, B \in \mathfrak{B}\}.$$

We obtain $\mathfrak{A}_1 \wedge \mathfrak{A}_2 \wedge \cdots \wedge \mathfrak{A}_n$ by extending the definition inductively to collections $\mathfrak{A}_1, \mathfrak{A}_2, \cdots, \mathfrak{A}_n$, for an arbitrary integer $n > 1$.

THEOREM. *If $\mathfrak{A}$ and $\mathfrak{B}$ are closed, closure preserving collections of subsets of a topological space, then $\mathfrak{A} \wedge \mathfrak{B}$ is closed and closure preserving.*

Proof. Let $\mathfrak{A} = \{A_i \mid i \in I\}$, $\mathfrak{B} = \{B_j \mid j \in J\}$ and let $R \subseteq I \times J$. To show $F = \bigcup\{A_i \cap B_j \mid (i, j) \in R\}$ is closed, it suffices to verify that each $x \notin F$ has a neighborhood not meeting F. Since $x \notin (A_i \cap B_j)$ for each $(i, j) \in R$ we have either $x \notin A_i$ or $x \notin B_j$. Let $p_1: I \times J \to I$ and $p_2: I \times J \to J$ be the projections, let $I' \subseteq p_1 R$ be those indices for which $x \notin A_i$ and let $J' \subseteq p_2 R$ be those indices for which $x \notin B_j$. Then the set $W = [Y - \bigcup\{A_i \mid i \in I'\}] \cap [Y - \bigcup\{B_j \mid j \in J'\}]$ is open, since the families $\mathfrak{A}$ and $\mathfrak{B}$ are closed and closure preserving, and W clearly contains x. Moreover, $W \cap F = \varnothing$: for, given $(A_i \cap B_j)$ with $(i, j) \in R$, either $i \in I'$ or $j \in J'$; if $i \in I'$ then $A_i \cap B_j \subseteq A_i$ whereas $W \subseteq Y - A_i$, and similarly, if $j \in J'$ then $A_i \cap B_j \subseteq B_j$ whereas $W \subseteq Y - B_j$.

We remark that every locally finite collection of subsets of a topological space is closure preserving but the converse is not true.

THEOREM. *A T_1 space X is metrizable if and only if there exists a sequence $\{\mathfrak{F}_n \mid n$ a positive integer$\}$ of closed, closure preserving coverings of X such that for any open set U and $x \in U$ there exists a positive integer n with $\bigcup\{F \in \mathfrak{F}_n \mid x \in F\} \subseteq U$.*

Proof. Simply define $\mathfrak{F}_n = \mathfrak{F}_1 \wedge \mathfrak{F}_2 \wedge \cdots \wedge \mathfrak{F}_n$, apply the above theorem, and without further modification carry out the proof as given in [**1**, p. 196].

References

1. J. Dugundji, Topology, Allyn and Bacon, Boston, 1966.

2. K. Morita, A condition for the metrizability of topological spaces and for n-dimensionality, Sci. Rep. Tokyo Kyoiku Daigaku Sect. A, 5 (1955) 33–36.

A PHYSICAL APPLICATION OF A REARRANGEMENT INEQUALITY

M. S. KLAMKIN, Ford Motor Company, Dearborn, Mich.

A usual problem in calculus and differential equations is to find the time of discharge of a tank with liquid through an orifice. Here the relevant differential equation is

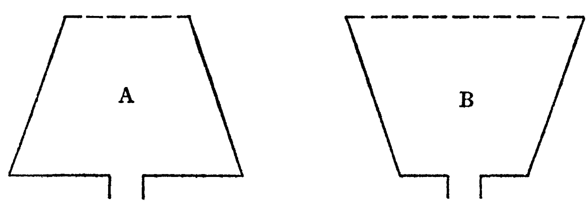
$$dV/dt = A(h)(dh/dt) = -cA_o\sqrt{2gh},$$

where

h = height of liquid above orifice,
$A(h)$ = cross-sectional area of tank at height h,
A_o = area of orifice,
g = gravitational constant,
c = coefficient of discharge; approximately
= .61 or $\pi/(\pi+2)$ for turbulent flow.

For a nonroutine problem in tank discharge, consider the following exercise [1]:

Shown below are schematic figures for two identical tanks (except for orientation) which are frustums of a right circular cone. If both tanks are initially full of water, which tank will empty out first? (Assume that the orifice areas are congruent and that the tanks start discharging at the same time.)



The usual method of solving this problem is to integrate the two relevant differential equations to find the times of discharge and then to compare times. A much more elegant solution can be gotten physically by noting that discharge rate for tank B is greater than the corresponding discharge rate for tank A when compared at equal cross-sections of B which are above its middle cross-section. This more than makes up for the low cross-sections when the discharge rate for tank B is less than that for tank A.

The above result will also hold for an arbitrary tank (and its inversion) whose cross-sectional area $A(h)$ is a monotonic decreasing function of the height above the orifice. Mathematically, this is equivalent to showing that

$$\int_0^H \frac{A(h)dh}{\sqrt{h}} > \int_0^H \frac{A(H-h)dh}{\sqrt{h}}.$$

(Note that here $A(H-h)$ is the area function for the inverted tank.)

The latter inequality will follow from the more general one:
If $D(x)$ and $E(x)$ are arbitrary monotonic decreasing functions, then

$$I = \int_0^{2a} D(x)E(x)dx > J = \int_0^{2a} D(2a-x)E(x)dx.$$

Proof. I and J can be rewritten as

$$I = \int_0^a \{D(x)E(x) + D(2a-x)E(2a-x)\}dx,$$

$$J = \int_0^a \{D(2a-x)E(x) + D(x)E(2a-x)\}dx.$$

Then,

$$I - J = \int_0^a \{D(x) - D(2a-x)\}\{E(x) - E(2a-x)\}dx > 0$$

since the integrand is nonnegative.

A discrete analog of the previous problem would be to arrange n persons at specified positions on one side of a see-saw to obtain the maximum turning moment. Physically, it is intuitive that the heaviest person should be at the end position, the next heaviest person at the next to end position, etc. A mathematical proof is given by the following rearrangement theorem for two sequences [2] and is analogous to the previous integral proof:

If $\{a\}$ and $\{b\}$ are two given sequences except for arrangement, then $\sum ab$ is greatest when $\{a\}$ and $\{b\}$ are monotonic in the same sense and least when they are monotonic in opposite senses.

Proof. Suppose that the a's are in ascending order but not the b's. Then there are a j and a k such that $a_j \leqq a_k$ and $b_j > b_k$. Since

$$a_jb_k + a_kb_j - (a_jb_j + a_kb_k) = (a_k - a_j)(b_j - b_k) \geqq 0,$$

we do not diminish $\sum ab$ by exchanging b_j and b_k. A finite number of such exchanges leads to an ascending order of the b's. The other half of the theorem is proved in the same way.

More general rearrangement theorems are given in [2, Chap. X].

References

1. R. E. Johnson and F. L. Kiokemeister, Calculus with Analytic Geometry, Allyn and Bacon, Boston, 1965, p. 340.

2. G. H. Hardy, J. E. Littlewood, and G. Pólya, Inequalities, Cambridge University Press, London, 1934, pp. 261–262.

MATHEMATICAL EDUCATION

EDITED BY J. G. HARVEY AND M. W. POWNALL

Material for this Department should be sent to either of the editors: J. G. Harvey, Department of Mathematics, University of Wisconsin, Madison, WI 53706; M. W. Pownall, Department of Mathematics, Colgate University, Hamilton, NY 13346.

GEOMETRY PREPARATION FOR HIGH SCHOOL MATHEMATICS TEACHERS

RUTH E. M. WONG, University of Hawaii

Introduction. The high school geometry course has been the subject of much debate in recent years and numerous recommendations have been made by individuals and professional groups. Accompanying these diverse proposals for modification in the high school geometry program have been concern for and indecision regarding the nature of adequate preparation in geometry for mathematics teachers.

As one aspect of a study on geometry for teachers [1], a survey of selected teacher education institutions was conducted during the academic year 1967–68, to ascertain the nature of geometry preparation provided and to obtain recommendations for the future.

Stratified sampling was used with the following three categories: state universities, teachers colleges, and other teacher education institutions. For purposes of this investigation, "state university" was defined as a state institution which offers graduate degrees at the doctoral level in liberal arts, general, and professional fields; "teachers college" meant not only an institution with "teachers" in its name, but any state college or institution which prior to 1963 was primarily concerned with teacher preparation as indicated by its accreditation in teacher education only; "other" included liberal arts colleges and private universities with general and professional programs. Only institutions accredited by nationwide or regional accrediting associations were considered for possible inclusion. Wherever possible at least one institution in each category was selected from each state.

Two forms of a questionnaire were developed, one each for the mathematics and education departments in each institution. Respondents were generally chairmen of their respective departments or individuals responsible for teacher training. Replies were received from at least one department in each of 155 institutions, making a 92.2 percent return from the institutions contacted, and from both departments in 130 institutions. Of the 130 institutions, 48 were state universities, 43 were teachers colleges, and 39 were other teacher education institutions. Since one or two questions were not answered on most forms, the total number of responses varied from item to item. Percents were calculated on the basis of total responses for a given item.

Geometry Courses for Teachers. Prospective secondary school mathematics teachers were required to take (or majority took) at least one course in geometry

in 93 percent of the 123 institutions reporting. In most cases, institutions offered more geometry courses than were required of prospective teachers. As shown in Table **1**, the number of required courses varied from zero to three while the number of available courses ranged from zero to five.

TABLE 1. Number of Institutions Offering Designated Number of Geometry Courses

	Number of Courses											
	0		1		2		3		4		5	
	f	%	f	%	f	%	f	%	f	%	f	%
a. Required of Teachers	8	7	62	50	44	36	9	7	0		0	
b. Available to Teachers	6	5	32	27	48	37	25	19	12	9	4	3

In a breakdown by type of institution, it appeared that state universities offered a wider range of geometry courses than either teachers colleges or liberal arts and private colleges, state universities being the only ones offering as many as five courses. This was not surprising since state universities, by the definition used in this study, offered the widest range of fields of study and generally had larger enrollments and faculties than the others. The few institutions which had no geometry courses available were among the 'other' teacher education institutions. Again, it would seem that the nature of their program, as in the case of small liberal arts colleges, influenced their restricted offering.

Table 2 shows the number of geometry courses available to prospective mathematics teachers by type of institution. The number of courses most commonly offered by both state universities and teachers colleges was two, available in approximately 43 percent of them. The same percentage of other institutions offered one course.

TABLE 2. Number of Geometry Courses Available to Prospective Mathematics Teachers

Number of Courses	Institutions						Total	
	State		Teachers		Other			
	f	%	f	%	f	%	f	%
0	0		0		6	4.6	6	4.6
1	5	3.8	11	8.5	19	14.6	35	26.9
2	19	14.6	19	14.6	10	7.8	48	36.9
3	11	8.5	9	6.9	5	3.8	25	19.2
4	5	3.8	4	3.1	3	2.3	12	9.2
5	4	3.1	0		0		4	3.1
Total	44	33.8	43	33.1	43	33.1	130	99.9

A classification of existing courses, based on specific textbooks reported and

course descriptions and outlines provided, was made as follows: "contemporary geometry" was defined as a course primarily concerned with axiomatic structure and a re-examination of elementary Euclidean geometry, "survey of geometry" as a course which treats several geometries, and "college geometry" as the traditional advanced Euclidean geometry. This admittedly rough classification revealed that of 121 institutions reporting, 66 percent offered contemporary geometry, 40 percent projective geometry, 34 percent survey of geometry, 12 percent college geometry, 11 percent non-Euclidean geometry, and 10 percent differential geometry.

Institutions were generally in favor of emphasizing transformations in college level geometry, 60 percent indicating this. The majority were also in favor of approaching transformations from algebraic and synthetic points of view.

The chi square statistic was used to test the independence of these responses from department and type of institution.

There was no significant difference between departments or among institutions in response to the question regarding emphasis on transformations in geometry for teachers. However, it may be noted that there were a higher percentage of 'yes' and lower percentages of 'no' and 'undecided' responses from mathematics departments than from education departments. More specifically, Table 3 shows that 68/100 or 68 percent of the mathematics departments advocated emphasis on transformations, while 49/95 or approximately 50 percent of the education departments favored it. Six out of 100 or 6 percent of the mathematics departments and 11/95 or approximately 12 percent of the education departments were not in favor of transformations, and 26/100 or 26 percent of the mathematics departments and 35/95 or approximately 36 percent of the education departments were undecided.

Among the institutions, state universities had the highest percentage of 'yes' and other institutions the highest percentage of 'undecided' responses.

TABLE 3. Emphasis on Transformations in College Geometry by Departments

Response	Departments				Total	
	Mathematics		Education			
	f	%	f	%	f	%
Yes	68	34.9	49	25.1	117	60.0
No	6	3.1	11	5.6	17	8.7
Undecided	26	13.3	35	17.9	61	31.3
Total	100	51.3	95	48.7	195	100.0

There were significant differences between responses from the two departments regarding algebraic and synthetic approaches to transformations at $P<.05$; i.e., chi square values as large as 8.847 for Table 4 and 6.793 for Table 5

could occur by random sampling alone less than five out of a hundred times. While the majority of each of the departments were in favor of an algebraic approach to transformations, mathematics departments gave a much higher 'yes' response and education departments a relatively high 'undecided' response. As shown in Table 4, 65 out of 91 or approximately 71 percent of the mathematics departments advocated an algebraic approach while 49/89 or approximately 55 percent of the education departments were in favor of it. 'Undecided' responses were given by 15/91 or approximately 16 percent of the mathematics departments in contrast to 32/98 or approximately 36 percent of the education departments.

TABLE 4. Transformations Developed from Algebraic Point of View by Departments

Response	Departments				Total	
	Mathematics		Education			
	f	%	f	%	f	%
Yes	65	36.1	49	27.2	114	63.3
No	11	6.1	8	4.4	19	10.6
Undecided	15	8.3	32	17.8	47	26.1
Total	91	50.6	89	49.4	180	100.0

The significant differences between departments on the synthetic approach to transformations were again due to a high percentage (50/83 or 60%) of 'yes' responses from mathematics departments and a higher percentage (33/81 or 41%) of 'undecided' responses from education departments as shown in Table 5. These differences may have been due to the division existing in many institutions between subject matter and professional training, with decisions as to subject matter left wholly to mathematics departments.

TABLE 5. Transformations in College Geometry Approached Synthetically by Departments

Response	Departments				Total	
	Mathematics		Education			
	f	%	f	%	f	%
Yes	50	30.5	34	20.7	84	51.2
No	14	8.5	14	8.5	28	17.1
Undecided	19	11.6	33	29.1	52	31.7
Total	83	50.6	81	49.4	164	100.0

There was no real difference among institutions on questions of algebraic and synthetic approaches to transformations. However, the highest percentage

of 'yes' responses was given by state universities and the highest percentage of 'undecided' by 'other' institutions.

Of 210 respondents, 62 percent were satisfied with the geometry courses being offered to prospective teachers in their institutions. Despite this high degree of satisfaction expressed, when asked for recommendations for change, only 20 percent recommended no change. Thirty-three percent felt that there was a need to change emphases in existing courses, 31 percent recommended adding courses to those presently available or requiring additional courses, 11 percent suggested both a change of emphasis and additional coursework, and 5 percent suggested other changes.

There were 134 institutions which responded to an invitation to comment on the teacher education program in geometry. Remarks made by eight or more individuals were as follows:

a. Satisfied with present program	22
b. Use variety of approaches to geometry; include study of various geometries	14
c. Teacher emphasis rather than content; teacher needs to acquire "spirit" of geometry, appreciation of mathematics	13
d. Relate geometry to rest of mathematics, to other subject areas	10
e. Stress axiomatic structure, logic, proof	10
f. Change emphasis from that of developing research mathematicians; relevance to high school	8
g. Increase requirements in geometry	8

Of those expressing satisfaction with their present program, 72.7 percent were from institutions offering a contemporary geometry course as described earlier. Mathematics departments advocated more strongly relating geometry to the rest of mathematics and to other subject matter areas, stressing axiomatic structure and proof, and increasing requirements. The greatest discrepancies among institutions occurred in the following ways: frequent mention by 'other' institutions of stress on axiomatic structure in contrast to one such response from state universities; a high frequency of suggestions for teacher emphasis rather than content only on the part of teachers colleges compared to a single similar suggestion from 'other' institutions; most recommendations for change of emphasis from that of developing research mathematicians from state universities in contrast to none from teachers colleges. Some of these differences may be a natural consequence of the program emphasized in each type of institution. For example, the training of research mathematicians may have dominated the graduate program in mathematics in state universities and may have caused concern on the part of those involved in teacher education that emphasis in undergraduate courses may not be appropriate for prospective teachers. On the other hand, teachers colleges whose primary focus has always been teacher preparation might have found this of less concern. Also, teachers colleges would probably have given greater consideration to aspects of the teacher preparation program beyond content of mathematics courses than 'other' institutions where the major emphasis is liberal arts.

Views on High School Geometry. Since teacher preparation in geometry

cannot be isolated from the high school geometry program, recommendations on secondary school geometry were also solicited from the same institutions.

There was no marked expression of satisfaction or dissatisfaction toward the high school geometry program as exemplified by SMSG geometry. Of the 201 responding to the question of whether they were satisfied, 35 percent said 'yes,' 30 percent 'no,' and 34 percent 'undecided.' There were no significant differences in response between departments or among institutions. This was the first item, however, in which there was a higher percentage of 'undecided' responses from mathematics departments than from education departments. This was probably due to the fact that the secondary schools and their programs are of major concern to those in education departments engaged in the training of high school teachers and hence there was more familiarity on their part with SMSG geometry and the benefits and problems of teaching it to high school students.

In analyzing responses by type of institution, it is interesting to note that while differences were very slight, the highest frequency among state universities was in a 'no' response, the highest among teachers colleges 'undecided,' and the highest among other institutions 'yes.'

Responses regarding various changes recommended in high school geometry are summarized in Table 6.

TABLE 6. Changes Recommended in High School Geometry

Change Advocated	Response						Total
	Yes		No		Undecided		
	f	%	f	%	f	%	
Less Emphasis on Formal Proof	58	31	99	53	30	16	187
More Coordinate Geometry	139	72	25	13	28	15	192
Vector Approach	73	40	41	22	71	38	185
Transformations Approach	82	44	37	20	68	36	187
More Rigor	41	22	97	53	45	25	183

There were significant differences between departments on the first two items in Table 6. The great majority (73%) of mathematics departments felt that there should not be less emphasis on formal proof whereas more education departments favored less emphasis than not as shown in Table 7. The differences were significant at the .001 level; hence they could have arisen by chance less than once in a thousand times.

Also, while coordinate geometry was favored by the group as a whole, a higher percentage of education departments advocated it and a higher percentage of mathematics departments were against it. The latter result was especially surprising in the light of the strong algebraic point of view expressed by mathematics departments earlier in relation to transformations in geometry courses

TABLE 7. Less Emphasis on Formal Proof Advocated for High School Geometry

Response	Departments				Total	
	Mathematics		Education			
	f	%	f	%	f	%
Yes	14	7.5	44	23.5	58	31.0
No	69	36.9	30	16.0	99	52.9
Undecided	11	5.9	19	10.2	30	16.0
Total	94	50.3	93	49.7	187	100.0

for teachers. The differences between departments shown in Table 8 were significant at the .01 level.

TABLE 8. More Coordinate Geometry Advocated for High School Geometry

Response	Departments				Total	
	Mathematics		Education			
	f	%	f	%	f	%
Yes	64	33.3	75	39.1	139	72.4
No	20	10.4	5	2.6	25	13.0
Undecided	12	6.3	16	8.3	28	14.6
Total	96	50.0	96	50.0	192	100.0

Reactions to questions relating to vector and transformations approaches to high school geometry were characterized by a relatively high percentage of 'undecided' responses. One possible explanation for the high 'undecided' response is that unlike coordinate geometry which was recommended as early as 1959 and has been incorporated in instructional materials for high school which have been available for several years, vector geometry and transformation geometry for high school can still be considered in the experimental stages with few materials readily available.

Institutions differed significantly ($P < .05$) in their response to more rigor. As shown in Table 9, the majority of state universities did not advocate more rigor, but less than half of the institutions in each of the other categories expressed similar views. More teachers colleges were in favor of more rigor than any of the other institutions and the high 'undecided' response from 'other' institutions was evident again.

In considering the question of more rigor in relation to the earlier one on less emphasis on formal proof, one might be tempted to think of them as opposites

and assume that the majority view against less emphasis on formal proof would imply support of more rigor in high school geometry. This was not the case. The majority did not advocate more rigor. It appeared that either respondents were making a distinction between emphasis on formal proof and rigor, or were saying that the present high school geometry program was satisfactory in terms of proof and rigor (neither more nor less) and that no change was recommended.

TABLE 9. More Rigor Advocated for High School Geometry by Type of Institution

Response	Institutions						Total	
	State		Teachers		Other			
	f	%	f	%	f	%	f	%
Yes	11	6.0	16	8.7	14	7.7	41	22.4
No	49	26.8	25	13.7	23	12.6	97	53.0
Undecided	14	7.7	11	6.0	20	10.9	45	24.6
Total	74	40.4	52	28.4	57	31.1	183	100.0

Summary. The high rate of return in the survey of colleges and universities was interpreted as an indication of interest in and concern for the problem of teacher preparation in geometry. At the same time the relatively high percentage of 'undecided' and omitted responses on many questions confirmed the general indecision indicated in the literature.

The survey revealed that present training of mathematics teachers at most of the selected institutions includes at least one required course in geometry, with Euclidean geometry from the contemporary viewpoint being the most frequent offering. While there seemed to be general satisfaction with teacher preparation in geometry, the majority hoped to see some change in the existing programs.

The inclusion of transformations in geometry for teachers was supported on the basis of strong recommendations from mathematicians and educators.

The high school geometry course desired by most university mathematicians and educators seemed to be one similar to SMSG geometry in degree of emphasis on formal proof and rigor but with more coordinate geometry incorporated.

Differences of opinion between mathematicians and educators may have been due to a number of factors, including lack of knowledge or appreciation of the situation. In any case, it must be recognized that these individuals occupy positions in their respective institutions where they can influence decisions on the teacher preparation program. Significant differences of opinion among such persons seem to suggest the need for careful and thoughtful consideration of the bases for all recommendations. They point to the advisability of continued collaboration among university mathematicians, educators, and competent

experienced teachers, not only in preparing instructional materials for high school geometry and determining the content of mathematics courses for teachers, but in examining the total preparation program of mathematics teachers. Each one—the mathematician, the mathematics educator, the high school teacher—has deep interest in, intimate knowledge of, and rich experience in his own specialized area, possessed by no one outside his field. Each should thus have a unique contribution to make in this determination.

Reference

1. Ruth E. M. Wong, Status and Direction of Geometry for Teachers, Ph.D. thesis, University of Michigan, 1968.

Erratum: In the article, "On the Ph.D. in Mathematics," by I. N. Herstein, on page 821, line 26, of the August-September 1969 issue of the MONTHLY, please read "damn" instead of "darn."

PROBLEMS AND SOLUTIONS

EDITED BY EMORY P. STARKE

All problems (both elementary and advanced) proposed for inclusion in this Department should be sent to E. P. Starke, 1000 Kensington Ave., Plainfield, NJ 07060. Proposers of problems are urged to enclose any solutions or information that will assist the editors. Ordinarily, problems in well-known textbooks and results in generally accessible sources are not appropriate for this Department. No solutions (except those accompanying proposals) should be sent to Professor Starke.

ELEMENTARY PROBLEMS

Solutions of Elementary Problems should be sent to Problems Group, Mathematics Department, University of Maine, Orono, ME 04473. To facilitate their consideration, solutions of Elementary Problems in this issue should be typed (with double spacing) on separate, signed sheets and should be mailed before April 30, 1970. Contributors (in the United States) who desire acknowledgement of receipt of their solutions are asked to enclose self-addressed stamped postcards.

E 2208. *Proposed by D. Rameswar Rao, Secunderabad, India*

Let $n \geqq 5$ and $2 \leqq b \leqq n$. Prove

$$\left[\frac{(n-1)}{b}\right] \equiv 0 \pmod{b-1}.$$

E 2209. *Proposed by M. S. Klamkin, Ford Scientific Laboratory*

Determine the locus of the centroids of all triangles similar to a given triangle and inscribed in another given triangle.

E 2210. *Proposed by R. N. Lloyd, University of Nottingham, England*

While the Dormouse lay sleeping in the middle of the table, the Mad Hatter and the March Hare sat down for tea. A number of places had been laid, and the Hatter and the Hare each moved alternately to a seat next to his old one. Each changed his direction of movement when, and only when, either the other had just changed his or else so as to avoid their both sitting on the same chair. Tea finished when both the Hatter and the Hare were again sitting in their initial positions. After how many moves was the tea-party over?

E 2211. *Proposed by M. D. Landau, Lafayette College*

An exercise in Rudin, *Principles of Mathematical Analysis*, states $\int_x^{x+1} \sin t^2 dt < 2/x$ for $x > 0$. Can this bound be improved to some smaller multiple of $1/x$?

E 2212. *Proposed by G. E. Peterson, University of Missouri, at St. Louis*

If $\{s_n\}$ and $\{t_n\}$ are sequences such that $\sum_{k=1}^{\infty} |t_k|$ converges and $\lim_{n\to\infty} (\sum_{k=1}^{n} s_k t_k)/s_n$ exists and is nonzero, then $\lim_{n\to\infty} s_n$ exists and is nonzero.

E 2213. *Proposed by H. Demir, Middle East Technical University, Ankara, Turkey*

Let us say that a (planar) polygon has the *Nagel property* if the lines through the vertices of the polygon and bisecting the perimeter of the polygon are concurrent. It is known that all triangles have the Nagel Property and that not all quadrilaterals have the property. Determine the simple nondegenerate quadrilaterals that have the Nagel property.

E 2214. *Proposed by M. S. Klamkin, Ford Scientific Laboratory and B. Ross Taylor, York High School*

It is intuitive that every simple n-gon ($n > 3$) possesses at least one interior diagonal. For a simple n-gon what is the least number of diagonals which, except for their endpoints, lie wholly in its interior?

SOLUTIONS OF ELEMENTARY PROBLEMS

A Necessary and Sufficient Condition

E 2143 [1969, 82, 413]. *Proposed by Peter Kornya, University of British Columbia*

In a triangle with sides a, b, c the line joining the centroid and the incenter is perpendicular to the bisector of the angle opposite side c. Show that the arithmetic mean of a, b, c equals the harmonic mean of a and b.

I. *Solution by F. Leuenberger, Feldmeilen, Switzerland.* Let G denote the centroid, I the incenter, Δ the area, h_a and h_b the altitudes to sides a and b, and r the inradius of triangle ABC. Let P and Q denote the points of intersection of line GI with sides BC and CA. Since triangles GPC and GQC have the same total area as triangles IPC and IQC, we have $\frac{1}{3}h_a+\frac{1}{3}h_b=2r$. The desired equation now follows immediately from

$$h_a=\frac{2\Delta}{a},\quad h_b=\frac{2\Delta}{b},\quad \text{and}\quad r=\frac{2\Delta}{a+b+c}\,.$$

II. *Solution by L. D. Goldstone, Watervliet, N. Y.* In trilinear normals, the line through $P(f, g, h)$ parallel to $l\alpha+m\beta+n\gamma=0$ is

$$\frac{lf+mg+nh}{l\alpha+m\beta+n\gamma}=\frac{af+bg+ch}{a\alpha+b\beta+c\gamma},$$

(cf. Smith, *Conic Sections*, p. 346). GI is parallel to the external bisector of angle C (for which $\alpha+\beta=0$), hence take $l=m=1$. $n=0$. Let $P\equiv I(1, 1, 1)$, $f=g=h=1$. Hence

$$\frac{1+1}{\alpha+\beta}=\frac{a+b+c}{a\alpha+b\beta+c\gamma}\,.$$

The line passes through $G(1/a, 1/b, 1/c)\equiv(\alpha, \beta, \gamma)$, which implies

$$\frac{1+1}{(1/a)+(1/b)}=\frac{a+b+c}{1+1+1}\,.$$

Also solved by Anders Bager (Denmark), Leon Bankoff, M. G. Greening (Australia), C. V. Subba Rama Iyer (India), Vinay G. Kane (India), Norman Miller, Simeon Reich (Israel), Sister Stephanie Sloyan, P. Naga Sundaram (India), C. S. Venkataraman (India), A. W. Walker, Charles Wexler, Mark Yu, and the proposer.

Polynomial Approximations on All of the Real Line

E 2165 [1969, 413]. *Proposed by J. L. Kazdan, University of Pennsylvania*

If $f\in C(R)$ can be uniformly approximated throughout R by polynomials, then f is itself a polynomial. Supply a proof, or a counterexample.

Solution by D. M. Bloom, Brooklyn College. The statement is true. If polynomials P_n approach f uniformly, then for some n we have $|P_i(x)-f(x)|<1$ for all $i\geqq n$, all x in R. Hence for all $i\geqq n$, $|P_i(x)-P_n(x)|<2$ (all real x), so that P_i-P_p is a bounded polynomial, and hence is constant. Taking limits as $i\to\infty$, $f-P_n$ is constant, and hence $f=P_n+C$ is a polynomial.

Also solved by S. M. Ambler, Joel Anderson, Anders Bager (Denmark), D. R. Brillinger, Charles Dunham, R. L. Enison, G. F. Feissner, D. E. Frohardt, D. S. Greenstein, D. W. Hadwin, D. A. Herrero, Ellen Hertzmark, G. A. Heuer, Kenneth Lang & Steven Minsker, Douglas Lind, Dan Marcus, M. D. Mavinkurve (India), Henrik Meyer (Denmark), Bryan Powers, Steve Rohde, R. A. Struble, Philip Trauber, J. E. Wilkins, Jr., Mark Yu, and the proposer.

Tossing a Polygon onto a Checkerboard

E 2166 [1969, 413]. *Proposed by Michael Stolnicki, Oakland Community College, Michigan*

Let a checkerboard consist of squares with sides of length 4. A regular $4n$-gon with radius 1 is tossed on the board. Determine the probability that the polygon will cross a line of the checkerboard.

Solution by Clyde Kessel, student, John Hersey High School, Prospect Heights, Ill. Without loss of generality we can assume the center of the polygon falls on or within one given square, at a distance d from one of the closest sides. Erect the perpendicular through the center of the polygon to the given side of the square. This perpendicular will form angles with the radii of the polygon. One of the angles with the smallest absolute value will be labelled θ. The vertex associated with this radius is the closest vertex to the given side since its distance is $d-\cos\theta$ and is a minimum when $\cos\theta$ is a maximum.

If the polygon crosses the given edge of the square d must be less than $\cos\theta$. Since the number of sides of the polygon is a multiple of 4, the situation with respect to all four sides of the square is equivalent. Therefore the center of the polygon must be within $\cos\theta$ units from a side of the square. The probability that this will occur is

$$(16\cos\theta - 4\cos^2\theta)/16 \tag{1}$$

This is the ratio of the area of the region in which the center may lie so that the polygon will intersect the square, to the area of the square.

$|\theta|$ must be less than $\pi/4n$. Therefore the probability that an angle close to any given θ will occur is

$$4n d\theta/\pi. \tag{2}$$

The probability that θ will occur and the polygon will intersect the square is the product of (1) and (2). The required probability is the sum of this product for all values of $|\theta|$, that is

$$\frac{4n}{\pi}\int_0^{\pi/4n}\left(\cos\theta - \frac{1}{4}\cos^2\theta\right)d\theta = \frac{4n}{\pi}\sin\frac{\pi}{4n} - \frac{n}{4\pi}\sin\frac{\pi}{2n} - \frac{1}{8}\cdot$$

Also solved by Slobodan Ćuk (Yugoslavia), Jordi Dou (Spain), Thomas Hughes, M. S. Klamkin, Harry Lass, and the proposer.

A Condition on a Subset of a Group

E 2168 [1969, 413]. *Proposed by Rosta János, Central Research Institute for Physics, Budapest, Hungary*

Let G denote a group and K a subset of G. Prove the following theorem: If for a given natural number $n \geqq 2$, the relations $K^{-1} \subseteq K$ and $K^n \subseteq K$ hold,

then K^{n-1} is a subgroup of G and K^{n-2} is identical with some (e.g., left) coset of K^{n-1} in G. (Here K^i denotes as usual the set of all products of the form $k_1k_2 \cdots k_i$ with $k_j \in K$, $j=1, 2, \cdots, i$.)

Solution by Douglas Lind, Stanford University If $k=k_1 \cdots k_{n-1} \in K^{n-1}$, then $k_i^{-1} \in K^{-1} \subseteq K$, so $k^{-1}=k_{n-1}^{-1} \cdots k_1^{-1} \in K^{n-1}$. Also, if $m=m_1 \cdots m_{n-1} \in K^{n-1}$, then $km_1 \in K^n \subseteq K$, so $km=(km_1)(m_2 \cdots m_{n-1}) \in KK^{n-2}=K^{n-1}$. Hence K^{n-1} is a subgroup of G.

Let e be the identity of G. Since $K^{-1} \subseteq K$, $e \in K^2$, so $K^{r-2} \subseteq K^r$ for all $r \geqq 2$. Hence if $n=2s$ is even, then $K^2 \subseteq K^4 \subseteq \cdots \subseteq K^{2s} \subseteq K$, which implies K is a subgroup of G, and the result is immediate except in the trivially false case $n=2$. If $n=2t+1$ is odd, then $K \subseteq K^3 \subseteq \cdots \subseteq K^{2t+1} \subseteq K$, so $K=K^3= \cdots =K^n$. Then for any $k \in K$, $kK^{n-1} \subseteq K^n=K^{n-2}$, while $K^{n-2}=k(k^{-1}K^{n-2}) \subseteq kK^{n-1}$, so $K^{n-2}=kK^{n-1}$ is a coset of K^{n-1}.

Also solved by Anders Bager (Denmark), W. N. Bell, Orin Chein, J. P. Comiskey, Ted Cullen Stuart DeSousa, Neal Felsinger, W. F. Fox, Robert Gilmer, M. G. Greening (Australia), D. W. Hadwin, Lee Hagglund, G. A. Heuer, H. S. Lieberman, Donna Martellotto & Kathy Shiple, J. V. Michalowicz, J. H. Oppenheim, E. F. Schmeichel, D. A. Sibley, R. C. Singal (India), Judith Soriano, Stephen Spindler, E. J. Taft, W. A. Thrash, Jr., Philip Trauber, and the proposer.

Sibley remarks that K^{n-1} is normal in the subgroup of G generated by K. That the second statement of the problem is meaningless for $n=2$ was pointed out by the proposer but inadvertently omitted in the printing. Most of the solvers, however, noted the omission.

A Convergent Series of Arcs

E 2169 [1969, 414]. *Proposed by J. J. Hirstein, Illinois State University*

Let U be a unit circle. Let C_1 be a circle whose diameter is half the diameter of U. Recursively, let C_{k+1} be a circle whose diameter is half the diameter of C_k. Put C_1 tangent to U on the inside (at a_1). Put C_2 tangent to C_1 on the outside and tangent to U on the inside (at a_2). Continue in this manner, putting C_n tangent to C_{n-1} on the outside and tangent to U on the inside (at a_n). Let s_i be the arc length $a_i\, a_{i+1}$. Does $\sum_{i=1}^{\infty} s_i$ converge? If so, what is the limit?

Solution by Simeon Reich, Israel Institute of Technology, Haifa. Let O be the center of U, and let O_n be the center of the circle C_n. Applying the law of cosines to the triangle $O_{n+1}O_nO$, we obtain

$$s_n = \arccos O_nOO_{n+1} = \arccos\left(1 - \frac{2}{(2^n - 1)(2^{n+1} - 1)}\right)$$
$$\leqq \arccos(1 - 1/2^{2(n-1)}) = t_n.$$

By l'Hospital's rule, $\lim_{n\to\infty} t_{n+1}/t_n = \frac{1}{2}$. By the ratio test, $\sum_{n=1}^{\infty} t_n$ converges. Since $s_n > 0$, the series in question also converges. Its sum is 2.04900825 approximately.

Also solved by Anders Bager (Denmark), Michael Goldberg, D. A. Herrero, G. A. Heuer, and Thomas Hughes.

Goldberg gives the sum as approximately $2\pi/3$, Hughes and this editor agree upon 2.0490114 with error less than 10^{-7}.

A Divergent Series

E 2170 [1969, 414]. *Proposed by M. Slater, University of Bristol, England*

Let $\{a_n\}$ be an increasing series of positive reals. Suppose $\lim_{n\to\infty} a_n = \infty$. Show that $\sum_{n=1}^{\infty} \cos^{-1}(a_n/a_{n+1}) = \infty$.

I *Solution by E. F. Schmeichel, The College of Wooster.* If $0<x_1<x_2<x_3$, it is readily verified that $\cos^{-1}(x_1/x_3) \leqq \cos^{-1}(x_1/x_2)+\cos^{-1}(x_2/x_3)$. (Simply take the cosine of both sides.) It follows easily by induction that if $0<x_1<x_2< \cdots <x_n$, then

$$\text{(1)} \qquad \cos^{-1}\left(\frac{x_1}{x_n}\right) \leqq \cos^{-1}\left(\frac{x_1}{x_2}\right) + \cos^{-1}\left(\frac{x_2}{x_3}\right) + \cdots + \cos^{-1}\left(\frac{x_{n-1}}{x_n}\right).$$

Set $k_0=1$. Let k_1 be the least positive integer such that $a_{k_1} \geqq 2a_{k_0}$, and in general, let k_j be the least positive integer such that $a_{k_j} \geqq 2a_{k_{j-1}}$. Note that $a_{k_{j-1}}/a_{k_j} \leqq \frac{1}{2}$.

If S_n denotes the nth partial sum of the given series, then by (1) we have

$$S_{k_n-1} = \sum_{i=1}^{k_n-1} \cos^{-1}\left(\frac{a_i}{a_{i+1}}\right) \geqq \sum_{j=1}^{n} \cos^{-1}\left(\frac{a_{k_{j-1}}}{a_{kj}}\right) \geqq n \cos^{-1}\left(\frac{1}{2}\right) > n.$$

So the partial sums of the series are unbounded, implying divergence as required.

II. *Solution by D. A. Zave, UNIVAC, Roseville, Minn.* Since the second derivative of $\cos^{-1}$ is negative on (0, 1), we observe that $\cos^{-1}$ is convex upward on [0, 1]. It follows that

$$\cos^{-1}(x) \geqq \tfrac{1}{2}\pi(1-x) \geqq 1-x, \qquad x \in [0, 1].$$

If m and n are positive integers with $m \leqq n$, then

$$\sum_{k=m}^{n} \cos^{-1}\left(\frac{a_k}{a_{k+1}}\right) \geqq \sum_{k=m}^{n}\left(1-\frac{a_k}{a_{k+1}}\right) \geqq \sum_{k=m}^{n} \frac{a_{k+1}-a_k}{a_{n+1}} = 1-\frac{a_m}{a_{n+1}} \cdot$$

Letting $n\to\infty$, we observe that the tail of the series is never less than 1. It follows that the series does not converge.

Also solved by Anders Bager (Denmark), D. E. Frohart, Leon Gerber, M. S. Klamkin, Douglas Lind Dan Marcus, M. D. Mavinkurve (India), Jernej Polajnar (Yugoslavia), Simeon Reich (Israel), R. A. Struble, J. E. Wilkins, Jr., and the proposer.

Note. Solution II shows that $\cos^{-1}$ may be replaced by any function f such that $f(x) \geqq 1-x$ on $(a, 1]$, where $0 \leqq a < 1$.

A Minimum Partition Problem

E 2171 [1969, 414]. *Proposed by Kenneth Jackman, Federal Electric Corporation, Fairbanks, Alaska*

Given N, what is the smallest W for which $B_1+B_2+\cdots+B_c=W$, and $B_1B_2\cdots B_c \geqq N$, with all B_k positive integers.

Note. The statement of the problem is ambiguous since it is not clear whether the integer c is fixed. Solutions were submitted for both cases.

I. (c fixed.) *Solution by David Zeitlin, Minneapolis, Minnesota.* From the arithmetic-geometric inequality, we have

$$\frac{W}{c} = \frac{B_1 + B_2 + \cdots + B_c}{c} \geqq \sqrt[c]{B_1 B_2 \cdots B_c} \geqq \sqrt[c]{N}.$$

Thus, $W = c\sqrt[c]{N}$, if integral; otherwise, $W = [c\sqrt[c]{N}] + 1$.

II. (c not fixed.) *Solution by M. S. Klamkin, Ford Scientific Laboratory.* The dual of this problem is to find the largest number which can be obtained as the product of positive integers whose sum is $\leqq S$. This problem was proposed by Leo Moser and solved by L. Carlitz [Problem 125, *Pi Mu Epsilon Journal*, Fall, 1961]. If $P(S)$ denotes the maximum product, it was shown that

$$P(S) = \begin{cases} 3^m & \text{if } S = 3m, \\ 4 \cdot 3^{m-1} & \text{if } S = 3m + 1, \\ 2 \cdot 3^m & \text{if } S = 3m + 2. \end{cases}$$

Here S is partitioned into as many 3's as possible.

It now follows immediately that if $P(S) + 1 \leqq N \leqq P(S+1)$, then $W_{\min} = S + 1$ (the corresponding partition is not unique in general).

Also solved by M. T. Bird, Slobodan Ćuk & Jernej Polajnar (Yugoslavia), Michael Goldberg, M. G. Greening, (Australia), G. A. Heuer, T. F. Hughes, Douglas Lind, Henrik Meyer (Denmark), Norman Miller, E. F. Schmeichel, C. S. Venkataraman (India), and the proposer.

ADVANCED PROBLEMS

All solutions of Advanced Problems should be sent to J. Barlaz, Rutgers—The State University, New Brunswick, NJ 08903. To facilitate their consideration, solutions of Advanced Problems in this issue should be typed (with double spacing) on separate signed sheets and should be mailed before April 30, 1970. Contributors (in the United States) who desire acknowledgement of receipt of their solutions are asked to enclose self-addressed stamped postcards.

5707. *Proposed by W. A. Vasconcelos, Rutgers—The State University*

Let R be an integral domain and G a finite group. Assume the characteristic of R does not divide $|G|$. Prove that each R-derivation of $R[G]$ is inner.

5708. *Proposed by C. W. Avery, San Jose State College*

Let K be a finite extension of the field k, complete in a non-Archimedean valuation. Let $\overline{K}$ and $\bar{k}$ denote the residue class fields. Problem 16, p. 129 of P. J. McCarthy, *Algebraic Extensions of Fields* asserts that K is separable over k if $\overline{K}$ is separable over $\bar{k}$. Disprove.

5709. *Proposed by W. A. J. Luxemburg, California Institute of Technology*

For all $x>0$, determine

$$\lim_{n\to\infty}\frac{1}{(\sqrt{\pi})^n}\int_{D_n(x)}\cdots\int \exp[-(x_1^2+\cdots+x_n^2)]dx_1\cdots dx_n,$$

where

$$D_n(x)=\left\{(x_1,\cdots,x_n):\left|\frac{x_1}{1}+\frac{x_2}{\sqrt{2}}+\cdots+\frac{x_n}{\sqrt{n}}\right|\leqq x\right\}.$$

5710. *Proposed by R. E. Shafer, Lawrence Radiation Laboratory, Livermore, California*

It is well known that

$$[R^2-2Rr\cos\theta+r^2]^{-\nu}=\sum_{n=0}^{\infty}\frac{r^n}{R^{n+2\nu}}C_n^{\nu}(\cos\theta),$$

$|r|<|R|$. $\mathrm{Re}(\nu)>-1$, $\mathrm{Re}(\nu)\neq 0$. Find the set of functions $F_n(r,R)$ independent of θ such that

$$[R^2-2Rr\cos\theta+r^2]^{-\mu}=\sum_{n=0}^{\infty}F_n(r,R)C_n^{\nu}(\cos\theta),\quad \mathrm{Re}(\mu)>-1.$$

5711. *Proposed by Dan Marcus, York University, Toronto*

Let $A=(a_{m,n})_{m,n=1}^{\infty}$ be an infinite matrix of nonzero integers such that for each m, the set of prime divisors of numbers in the mth row is finite. Prove that the system of congruences $x_{m+n}\equiv a_{m,n}(\mathrm{mod}\ x_m)$ is solvable in primes.

5712. *Proposed by Dan Marcus, York University, Toronto*

Is it possible to topologize the integers in such a way that the connected sets are the sets of consecutive integers? Generalize to the lattice points of n-space.

5713. *Proposed by D. P. Geller, University of Michigan*

For any graph G with p points, q lines, and chromatic number χ, show

$$\chi\geqq p^2/(p^2-2q).$$

SOLUTIONS OF ADVANCED PROBLEMS

Multiplicative Solutions of Some Number Theoretic Equations

5653 [1969, 200]. *Proposed by Richard Stanley, Harvard University*

Find in each case the real-valued multiplicative number-theoretic function f which satisfies the stated condition. μ is the Möbius function, ϕ is the Euler totient function, $d(n)$ is the number of divisors of n, and $\sigma(n)$ is the sum of the

divisors of n.

(A) $$\sum_{d|n} f(d) = \mu(n)f(n).$$ (C) $$\sum_{e|n} \phi(e)f(n/e) = d(n)f(n).$$

(B) $$|f(n)| = \sum_{d|n} f(d).$$ (D) $$\sum_{d|n} \frac{\mu(d)d^2f(n/d)}{\phi(d)} = \sigma(n)f(n).$$

Solution by M. S. Demos, Villanova University. Since f is multiplicative, it is enough to find $f(p^a)$ where p is a prime. In all cases $f(1)$ is arbitrary, except as noted in case (B). If we set in (A) and (B) $n=1, p, p^2, \cdots, p^a$ in succession, we obtain easily:

(A) $f(1) = c,\ \ f(p) = f(p^2) = -c/2,\ \ f(p^a) = 0\ \ \text{for } a \geqq 3.$

(B) $f(1) = c \geqq 0,\ \ f(p^a) = -c/2^a\ \ \text{for } a \geqq 1.$

(C) Setting $n = p^a$, the equation becomes

$$(a+1)f(p^a) = \sum_{j=0}^{a} \phi(p^j)f(p^{a-j}) = f(p^a) + (p-1)f(p^{a-1}) + \sum_{j=2}^{a} \phi(p^j)f(p^{a-j}).$$

We replace j by $j+1$, and notice that $\phi(p^{j+1}) = p\phi(p^j)$ for $j \geqq 1$, thus obtaining

$$\begin{aligned} af(p^a) &= (p-1)f(p^{a-1}) + p\sum_{j=1}^{a-1} \phi(p^j)f(p^{a-1-j}) \\ &= (p-1)f(p^{a-1}) + p\{af(p^{a-1}) - f(p^{a-1})\}. \end{aligned}$$

Then $f(p^a) = (ap-1)f(p^{a-1})/a$, and for $a \geqq 1$

$$f(1) = c,\quad f(p^a) = \frac{c(p-1)(2p-1)\cdots(ap-1)}{a!}.$$

(D) Set $n = p^a$. Then

$$f(p^a) - \frac{p^2f(p^{a-1})}{p-1} = \frac{p^{a+1}-1}{p-1}f(p^a),$$

whence $f(p^a) = -pf(p^{a-1})/(p^a-1)$. Finally for $a \geqq 1$,

$$f(1) = c,\quad f(p^a) = \frac{(-1)^a cp^a}{(p-1)(p^2-1)\cdots(p^a-1)}.$$

Also solved by T. M. Apostol, Anders Bager (Denmark), Joel Berman, W. J. Blundon, L. Carlitz, M. D. Chowdhury, Josef Daneš (Czechoslovakia), K. J. Davis, Ray Glenn, M. G. Greening (Australia), Emil Grosswald, O. P. Lossers (Netherlands), E. J. F. Primrose (England), Simeon Reich (Israel), F. G. Schmitt, Jr., A. P. Shah (India), D. A. Smith, Al Somayajulu, Philip Trauber, E. W. Trost (Switzerland), C. S. Venkataraman (India), and the proposer.

The proposer establishes the solution for all parts of the problem by obtaining generating functions similar to the methods used in solving 5293 [1966, 555] and 5446 [1967, 1274].

Application of a Lemma of Gauss in a Sum of Sines

5656 [1969, 200]. *Proposed by P. A. Catlin, Carnegie-Mellon University*

Let a and n be positive integers such that $2n+1$ is prime, and let the brackets denote the greatest integer function. When is $\sum_{j=1}^{n} [\sin(2\pi aj/(2n+1))]$ an even integer?

Solution by Joel Berman, University of Washington. Put $S=\sin(2\pi aj/(2n+1))$. If a and $2n+1$ are not relatively prime, then $2n+1$ divides a, $S=0$, and $\sum [S]=0$, which is even.

If $(a, 2n+1)=1$, then $[S]=0$ if $aj\equiv m \pmod{2n+1}$ where $0\leqq m\leqq n$; and $[S]=-1$ if $aj\equiv m \pmod{2n+1}$ where $n<m<2n+1$. Hence $\sum[S]$ is even if and only if the number v of values of j, $1\leqq j\leqq n$, for which aj is congruent to numbers greater than n and less than $2n+1$ is even. By a lemma of Gauss (cf. Niven and Zuckerman, *An Introduction to the Theory of Numbers*, p. 70), it follows that $a/(2n+1)=(-1)^v$, so that v is even if and only if a is a quadratic residue modulo $2n+1$.

Also solved by Merrill Barnebey, Robert Breusch, Arthur Gittleman, M. G. Greening (Australia), Henrik Meyer (Denmark), A. P. Shah & C. G. Khatri (India), Joel Spencer, and the proposer.

REVIEWS

EDITED BY KENNETH O. MAY, University of Toronto

COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER, Carleton College

Printed materials for review should be sent to Kenneth O. May, American Mathematical Monthly, Department of Mathematics, University of Toronto, Toronto 5, Canada. Films and correspondence relating to films should be sent to Seymour Schuster, Carleton College, Northfield, Minnesota 55057.

All unsigned material is written by the editors. A boldface capital **C** *in the margin indicates that a review is based in part on classroom use. Professors willing to write such a review should first inform the editor in order to avoid duplication.*

Note to Publishers: *Although the copyright law does not require permission for short quotations, we request that publishers wishing to quote from Reviews request permission from the author of the review and check the copy with him. For telegraphic reviews correspondence should be addressed to the editor of this section.*

Introduction to Complex Analysis. By Rolf Nevanlinna and V. Paatero. Translated by T. Kövari and G. S. Goodman. Addison-Wesley, Reading, Mass., 1969. ix+348 pp. $11.50. (Telegraphic Review, August/September 1969.)

A book on complex analysis having Rolf Nevanlinna as one of the authors has to be treated with respect if not veneration. It is one of the masters of the craft that speaks, one to whom the most spectacular advance in complex function theory of this century is due. What is said is said with great precision, and it is well said. One could only wish that more were said.

This is a book that should be at the beck and call of every mathematician who ventures to give a course in the theory of functions of a complex variable. It is full of novel points of view, alternate modes of attack, and 320 excellent problems. A knowledgeable teacher with a good class can use it as a text, and if he knows the why and wherefore of the various steps, he has here the basis for a very stimulating course. But he has to be prepared to provide the missing prose of a descriptive and motivational nature. A somewhat sophisticated student may use it for self-study. He must not be repelled by the terseness of the style, and he should be willing to work all the problems. This will give him a superb grasp of the subject and the satisfaction of having mastered a rigorous uptodate introduction to analytic function theory.

The book is singularly autonomous (a reference to a proof of the Jordan curve theorem is about the only concession). Historical information is doled out sparingly; only in the chapters on the Riemann zeta function and conformal mapping is the allowance more liberal. There is no reference to Rolf Nevanlinna in the text; for American readers this is a deplorable omission.

The reviewer would have liked a fuller treatment of the graph theoretical description of Riemann surfaces in Chapters 6 and 12. This is one of the many tools used by the "Finnish school" in the profound investigations of the value distribution problem for meromorphic functions, and here is one of the places where more descriptive prose would fill a need.

Section 10.5 has the title "Laplace Integrals." Granted that Laplace may have studied them; surely they are not what most mathematicians would understand by a Laplace integral. A word of caution to the reader would have been in order.

The elegant introduction of complex numbers in Chapter 1 as a two dimensional division algebra appears to be due to Weierstrass of all people. The use of homotopy theory in complex integration is welcome. So is the extensive use of harmonic functions and harmonic measure and many other innovations.

EINAR HILLE, University of New Mexico

C *Calculus, Volume 1 (One Variable Calculus, with an Introduction to Linear Algebra)*, 2nd ed. By Tom. M. Apostol, Blaisdell, Waltham, Mass., 1967.

This classic book, one of the first in the new wave of calculus texts which broke several years ago, has been improved in its new edition, with " . . . smaller chapters, . . . mean value theorem and routine applications . . . introduced . . . earlier, . . . new illustrative examples, . . . (expanded) applications to physics and engineering, and *many new and easier exercises*." (Emphasis mine.) It is now easier to use in the classroom and no less stimulating. A prospective user should consult both editions and the review of the first (this MONTHLY, 69 (1962) pp. 449–451) for information supplementing my comments, which are based primarily on my unorthodox but successful use of the book as a text for a second year course for freshmen and sophomores with varied backgrounds. I taught about $\frac{1}{3}$ of the course with no text.

I began with polynomial approximations (Chapter 7). This discussion of Taylor polynomials and error approximation precedes the introduction of sequences and series (Chapters 11 and 12), a splendid pedagogical device which stresses the importance of the remainder and the techniques required to estimate it. Similar careful preparation for the later introduction of sophisticated subjects is evident in the treatment of differential equations (Chapter 8). The stress on existence and uniqueness theorems in less than full generality and the numerous examples were both a good review and a base to which I could easily anchor the more advanced material appropriate to a second year course.

This edition contains a thorough and useful introduction to vector algebra and the study of curves as vector valued functions (Chapters 12–16) which was more than sufficient to prepare the students for my subsequent introduction of the Frechet derivative for scalar valued functions of a vector and a geometric discussion of line integrals and exact vector fields.

I could comfortably assign sections to students to read. There are no formulae for cranking out answers to the problems. The arguments are correct, clear, and analytical, though they rely perhaps a little too heavily on carefully designed notation and previously proved theorems. Students with no "mathematical maturity" might find them very hard going. I felt free—and compelled—to devote my lectures to the intuitive and geometric content of the material.

I thoroughly enjoyed using this book, although I did not "follow it." (Nor can I imagine one I would follow.) I recommend it for an honors class of freshmen or for students with some familiarity with college mathematics, for example, advanced placement students or high school teachers at a summer institute. All instructors should own a copy to use as a reference work. Let us profit from the author's labors.

E. D. BOLKER, Bryn Mawr College

An Introduction to Real Analysis. By Burton Randol. Harcourt, Brace & World, New York, 1969. xi+112 pp. $6.50. (Telegraphic Review, Aug./Sept. 1969.)

A popular approach to analysis involves a two semester course in elementary calculus followed by a two semester course in advanced calculus or real analysis with considerable attention given to modern aspects of topology. This approach has the mathematics major as its target. The non-math major usually takes one or two applied math courses to complete his mathematics minor.

The present text could be used effectively as a basis for one-semester course following elementary calculus. It is written with a clear and concise style. The topics considered are few in number but are prudently selected so as to allow maximum flexibility in designing a course which is student oriented. Its theorem-proof-remark format is particularly well-suited for such a course. The text is designed to continue and amplify the students work in elementary calculus. This text will occupy an important place in the analysis curriculum.

EDWARD PERESSINI, College of Great Falls, Montana

THE AMERICAN MATHEMATICAL MONTHLY

(FOUNDED IN 1894 BY BENJAMIN F. FINKEL)

THE OFFICIAL JOURNAL OF

THE MATHEMATICAL ASSOCIATION OF AMERICA, INC.

VOLUME 77 NUMBER 2

CONTENTS

(*continued on inside cover*)

FEBRUARY 1970

NOTICE TO AUTHORS

Articles should be typewritten on good quality paper, triple spaced with wide margins. Submit the original and a duplicate; keep a complete copy for protection against loss. Please follow the format used in current issues of the MONTHLY. The manuscript must make the author's intent clear to the printer. A paper accepted for publication by the editor but not acceptable to the printer will be returned to the author for retyping. Please use the American Mathematical Society *Manual for Authors of Mathematical Papers* as your guide and please conform to correct English usage. A *Manual for Monthly Authors* is available on request.

Prospectively MONTHLY authors are advised to consult the Statement of Policy in the January 1969 issue, p. 2. Specialized research is usually unsuitable. Backlog: Main Articles 11 months, Math. Notes 10 months, Research Problems 6 months, Classroom Notes 6 months, Math. Education 6 months.

EDITORIAL CORRESPONDENCE: to HARLEY FLANDERS, American Mathematical Monthly, Purdue University, Lafayette, IN 47907; ADVERTISING CORRESPONDENCE: to RAOUL HAILPERN, Mathematical Association of America, SUNY at Buffalo, Buffalo, N. Y. 14214; CHANGE OF ADDRESS and SUBSCRIPTIONS: to A. B. WILLCOX, Mathematical Association of America, 1225 Connecticut Ave., N.W., Washington, D.C. 20036.

Annual dues for members of the Association (including a subscription to the American Mathematical Monthly) are $10.00. For nonmembers the subscription price is $15.00.

PUBLISHED BY THE ASSOCIATION at Washington, D. C., and Menasha, Wisconsin during the months of January, February, March, April, May, June–July, August–September, October, November, December.

Second-class postage paid at Washington, D. C., and additional mailing offices. Acceptance for mailing at special rate of postage provided for in the Act of February 28, 1925, embodied in Paragraph 4, Section 538, P.L. and R., authorized April 1, 1926.

PRINTED IN THE UNITED STATES OF AMERICA

SOME ASPECTS OF FOURIER SERIES

CASPER GOFFMAN, Purdue University
DANIEL WATERMAN, Syracuse University

A continuous function whose Fourier series diverges may be "corrected" so that the Fourier series of the new function converges uniformly, either by an appropriate change of variable or by a modification of the given function on a suitable set of arbitrarily small measure.

Our purpose is to give an exposition of some mathematics which is more or less related to these phenomena. Certain of the matters we discuss seem to be ready for a more thorough treatment than they have as yet been given in the literature.

1. Preliminaries. We first remind the reader that a continuous function may have a Fourier series which diverges at some points. Let f be a continuous function of period 2π. Its Fourier series is given by

$$\frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nx + b_n \sin nx,$$

where

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t) \cos ntdt, \qquad b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t) \sin ntdt, \qquad n = 0, 1, 2, \cdots .$$

Then, if

$$s_n(x) = \frac{a_0}{2} + \sum_{k=1}^{n} (a_k \cos kt + b_k \sin kt), \qquad n = 0, 1, 2, \cdots,$$

each s_n has the form $s_n(x) = \int_{-\pi}^{\pi} D_n(x-t)f(t)dt$, where the functions D_n, $n=0, 1, 2, \cdots$, are called the *Dirichlet kernels*. The partial sums s_n of the Fourier series of f are thus the convolutions $s_n = D_n * f$.

If, instead of the partial sums s_n we consider their averages,

$$\sigma_n = \frac{1}{n+1}(s_0 + s_1 + \cdots + s_n),$$

Casper Goffman received his PhD at Ohio State Univ. under Henry Blumberg. He has held positions at Westinghouse, Univ. Kentucky, Univ. Oklahoma, Wayne, and Purdue, and has spent leaves at the Inst. Adv. Study and Westfield College, London. He has published extensively in real analysis and authored several books: *Real Functions*, *First Course in Functional Analysis* (with G. Pedrick), *Calculus of Several Variables*, and *Introduction to Real Analysis*.

Daniel Waterman wrote his Chicago thesis under Antoni Zygmund and has continued research in Fourier analysis and related fields. He has held positions with the Cowles Commission and at Purdue Univ., Univ. of Wisconsin, Milwaukee, Wayne State, and Syracuse. He spent a Fulbright year at the Univ. of Vienna and a year on leave at Berkeley. *Editor.*

we obtain

$$\sigma_n = \frac{D_0 + D_1 + \cdots + D_n}{n+1} * f = K_n * f.$$

The functions K_n are called the *Fejér kernels.* They are positive, whereas the D_n are oscillatory. For each $n = 0, 1, 2, \cdots$,

$$\int_{-\pi}^{\pi} D_n(t)dt = \int_{-\pi}^{\pi} K_n(t)dt = 1.$$

However, $\int_{-\pi}^{\pi} |D_n(t)| dt$ goes to infinity as n increases. Since the K_n are positive, we have $\int_{-\pi}^{\pi} |K_n(t)| dt = 1$. Moreover, it is easy to see that $K_n(t)$ converges uniformly to zero on $|t| > \delta$, for each $\delta > 0$. This implies that σ_n converges uniformly to f, for each continuous f.

We indicate why the fact that $L_n = \int_{-\pi}^{\pi} |D_n(t)| dt$ goes to infinity as n increases implies that there is a continuous f whose Fourier series diverges. We work at $x = 0$. Then, let

$$\phi_n(f) = s_n(0) = D_n * f(0).$$

For each n, ϕ_n is a linear functional on the space of periodic continuous functions of period 2π. A simple computation shows that L_n is the norm of ϕ_n. By the Banach-Steinhaus theorem, the unboundedness of $\{L_n\}$ implies that there is an f for which $\{\phi_n(f)\}$ is unbounded. The Fourier series of f is accordingly divergent at $x = 0$. The numbers L_n are called *Lebesgue constants.* Indeed, the above argument of elementary functional analysis was introduced by Lebesgue [**L**] in 1908.

2. A theorem of Pál and Bohr. Although there are continuous functions whose Fourier series do not converge, the situation may be amended by a simple change of variable as shown by the following result of Pál [**P**] and Bohr [**B**]:

THEOREM 1. *If f is a continuous function on $[-\pi, \pi]$, with $f(-\pi) = f(\pi)$, there is a homeomorphism g of $[-\pi, \pi]$ with itself such that the Fourier series of $f \circ g$ converges uniformly.*

The only known proof of this fact is accomplished by complex variable methods. We indicate the salient features of the proof, which involve a quite remarkable theorem of Fejér. The Riemann mapping theorem asserts that if D is a simply connected open set, not the whole plane, there is a conformal one-one mapping f of the unit disk onto D. Moreover, if D is a bounded set in the plane and its boundary is a simple Jordan curve, Carathéodory has shown that f may be extended to a homeomorphism between the closure of the unit disk and the closure of D. The theorem of Fejér asserts that if $\sum_{n=0}^{\infty} a_n z^n$ is the power series expansion of $f(z)$ in $|z| < 1$, then this series converges uniformly to the extension of f on $|z| \leq 1$. The proof of this theorem of Fejér is elementary. It involves the simple fact that the area of D is given by $\pi \sum_{k=1}^{\infty} k |a_k|^2$ together

with the following Tauberian theorem (which is also easy to prove): *If the series* $\sum a_k e^{ikx}$ *is uniformly* $(C, 1)$ *summable and* $\sum k|a_k|^2 < \infty$, *then the series converges uniformly.*

The theorem of Pál and Bohr follows from that of Fejér. By adding a continuous periodic function of bounded variation, the given function may be changed into a function f for which $f(-\pi)=f(\pi)=f(\xi)$ for exactly one $\xi \in (-\pi, \pi)$, unless f is identically a constant. It suffices to prove the theorem for such an f. Now let h be continuous and periodic, increasing on $(-\pi, \xi)$ and decreasing on (ξ, π). The mapping of $[-\pi, \pi]$ into the plane given by $z=h(t)+if(t)$ defines a simple closed curve. Let $w=F(z)$ map $|z|<1$ onto the domain bounded by this curve. F extends to a homeomorphism on $|z| \leqq 1$. Thus we have the function $F(e^{i\theta})$ mapping $[-\pi, \pi]$ one-one onto the curve $z=h(t)+if(t)$. We may select F so that $F(-1)=z(-\pi)$. Then if we let

$$g(t) = z^{-1} \circ F(e^{it}),$$

we have the theorem of Pál and Bohr.

It would be of interest to obtain a real variables proof of this theorem. In particular, extensions to other kinds of series might then become possible. Another question concerns the same phenomenon for complex valued functions. This amounts to the existence of a single homeomorphism g for which $f_1 \circ g$ and $f_2 \circ g$ have uniformly convergent Fourier series, for any two continuous real functions with $f_1(-\pi)=f_1(\pi)$ and $f_2(-\pi)=f_2(\pi)$. Of course, this question may be asked for n functions.

The recent theorem of Carleson, which says that if $f \in L_2$ then its Fourier series converges almost everywhere, implies that, for every finite-valued measurable f, there is a homeomorphism g such that the Fourier series of $f \circ g$ converges almost everywhere, since g may be taken so that $f \circ g \in L_2$. There should, however, be a direct proof of this fact.

3. The Salem criterion. Related to the above problem is that of the existence of continuous functions whose Fourier series converge uniformly after every change of variable. In this context, a criterion of Salem [S] on the uniform convergence of Fourier series is useful.

THEOREM 2. *If f is continuous, $f(-\pi)=f(\pi)$, and if the sum*

$$T_n(x) = \frac{f(x) - f\left(x + \frac{\pi}{n}\right)}{1} + \frac{f\left(x + \frac{2\pi}{n}\right) - f\left(x + \frac{3\pi}{n}\right)}{2} + \cdots + \frac{f\left(x + \frac{(n-1)\pi}{n}\right) - f(x+\pi)}{\frac{1}{2}(n+1)}$$

(n odd) as well as the corresponding sum obtained by substituting $-\pi$ for π, converge uniformly to zero as n increases, the Fourier series of f converges uniformly.

The proof of this theorem involves rather tedious but quite natural modifications of the condition that

$$\int_0^\pi \{f(x+t)+f(x-t)-2f(x)\}\,\frac{\sin nt}{t}\,dt,$$

converges uniformly to zero if and only if the Fourier series of f converges uniformly.

This criterion has interesting consequences. First, if ω is the modulus of continuity of f there is an absolute constant C such that

$$|T_n(x)| < C\log n\cdot\omega(\pi/n).$$

It follows that if $\omega(\pi/n)=o(1/\log n)$, then the Fourier series of f converges uniformly. This is the Dini-Lipschitz test.

As a second application of the Salem criterion we obtain the standard Jordan condition. Clearly,

$$|T_n(x)| < \omega\left(\frac{\pi}{n}\right)\left(1+\frac{1}{2}+\cdots+\frac{1}{m}\right)+\frac{V}{m+1},$$

where ω is the modulus of continuity of f, V is the variation of f, and m is any integer smaller than $\frac{1}{2}(n+1)$. If we choose m, as we may, so that m goes to infinity, and $\omega(\pi/n)\log m$ goes to zero, as n increases, we obtain the fact that if f is continuous, of bounded variation, and $f(-\pi)=f(\pi)$, then the Fourier series of f converges uniformly.

4. The Garsia-Sawyer criterion. An interesting generalization of this last condition also follows easily from the Salem criterion. This involves functions complementary in the sense of W. H. Young (see e.g., [**Zy**]), which we now define.

Let $\phi(u)$, $u\geqq 0$, be a continuous, strictly increasing function such that $\phi(0)=0$, and $\lim_{u\to\infty}\phi(u)=\infty$, and let $\psi(v)$ be the inverse of ϕ. Now, let

$$\Phi(u)=\int_0^u\phi(t)dt,\qquad \Psi(v)=\int_0^v\psi(t)dt,\qquad u,v\geqq 0.$$

The functions Φ and Ψ are said to be *complementary in the sense of Young.* We have Young's inequality: If $a\geqq 0$, $b\geqq 0$, then

$$ab\leqq\Phi(a)+\Psi(b).$$

By an elaboration of the above method, the following generalization of the Jordan condition is obtained. First, we say that a function f is of *bounded Φ-variation* if

$$\sup\sum_{i=1}^{n-1}\Phi(|f(x_{i+1})-f(x_i)|)<M$$

for some $M>0$, the sup taken over all partitions

$$-\pi = x_1 < x_2 < \cdots < x_n = \pi.$$

THEOREM 3. *If f is continuous, of bounded Φ-variation, and the complementary function Ψ of Φ is such that $\sum_{n=1}^{\infty} \Psi(1/n) < \infty$, then the Fourier series of f converges uniformly.*

The special case $\Phi(u) = u^p$, $p>1$, was given by L. C. Young [Y] before the work of Salem.

A number of questions remain. For example, if Φ is such that its complementary function Ψ does not satisfy the above condition, is there an f of bounded Φ-variation whose Fourier series does not converge uniformly?

A result on uniform convergence of Fourier series involving a different generalization of bounded variation has recently been given by Garsia and Sawyer [GS]. They start with the fact that a continuous f is of bounded variation if and only if

$$\int_0^1 n(y)dy < \infty,$$

where we assume $[0, 1]$ is the exact range of f, and $n(y)$ is the cardinality of $f^{-1}(y)$ when this is finite and ∞ otherwise. They make a slight change in the form of this condition. For each y, let $E_y = \{x : f(x) > y\}$. Then E_y is open. Let $N(y)$ be the number of components of E_y, considered modulo 2π. Then f is of bounded variation if and only if

$$\int_0^1 N(y)dy < \infty.$$

(We note that this is the one-dimensional case of an n-dimensional result of Fleming and Rishel [FR], where the role of $N(y)$ is taken by the perimeter of a set in the sense of de Giorgi.)

Garsia and Sawyer suppose $N(y)$ to be finite almost everywhere, and obtain an interesting decomposition of f. For each n and x, they let $f_n(x)$ be the measure of the set of points y, in the ordinate set, $[0, f(x)]$, of f at x, for which $N(y) = n$. Then $f_n(x)$ is continuous, and by considering integral means, is easily seen to be of bounded variation. Obviously, $f = \sum_{n=1}^{\infty} f_n$. By applying to this decomposition the fact that if E is an open set, with n components, then all the partial sums of the Fourier series of the characteristic function of E are bounded by

$$A + B \log n,$$

where A and B are absolute constants, they obtain the following result:

THEOREM 4. *If f is continuous, with $f(-\pi) = f(\pi)$, and such that*

$$\int_0^1 \log N(y)dy < \infty,$$

then the Fourier series of f converges uniformly.

It would be interesting to know if this result is best possible, and what connection there is between conditions such as those of Theorem 3 and those of Theorem 4.

5. Analogues for other groups. Since the behavior of the Walsh-Fourier series parallels that of the trigonometric Fourier series in many ways, it is natural to ask if Theorems 3 and 4 have analogues for Walsh series.

Suppose $x \in [0, 1]$ with dyadic expansion

$$x = \sum_{n=1}^{\infty} \frac{1}{2^n} d_n(x),$$

where $d_n(x) = 0$ or 1. We assume that the terminating expansion is chosen for dyadic rationals. Let

$$\psi_0(x) \equiv 1, \qquad \psi_{2^n}(x) = e^{i\pi d_n(x)}$$

and for $n = 2^{n_1} + 2^{n_2} + \cdots + 2^{n_k}$, where $0 \leqq n_1 < \cdots < n_k$ and

$$\psi_n(x) = \psi_{2^{n_1}}(x)\psi_{2^{n_2}}(x) \cdots \psi_{2^{n_k}}(x).$$

These are the *Walsh functions.* The functions $r_n = \psi_{2^n}$ are the *Rademacher functions.* The Walsh functions form a complete orthonormal system in $L^2[0, 1]$. If $f \in L^1[0, 1]$, we will denote the nth partial sum of its Walsh-Fourier series by $s_n(x, f)$.

The Walsh functions may be regarded as the characters of the direct product of a countable infinity of cyclic groups of order 2, 2^ω. In his now classical paper on Walsh-Fourier series, Fine [**F**$_1$] restructured the interval [0, 1] by introducing the topology and addition of the group 2^ω. Following his terminology, we define

$$x \dotplus x' = \sum \frac{1}{2^n} \{(d_n(x) + d_n(x')) \bmod 2\}.$$

The topology is determined by the metric

$$\rho(x, x') = \sum \frac{1}{2^n} \left| d_n(x) - d_n(x') \right|.$$

The following analogue of Theorem 2 has been demonstrated by C. Onneweer [**O**$_1$]:

THEOREM 5. *Let f be a continuous function of period* 1. *Let*

$$U_n(x) = \left|\frac{f(x \dotplus 2/2^{n+1}) - f(x \dotplus 3/2^{n+1})}{1}\right| + \left|\frac{f(x \dotplus 4/2^{n+1}) - f(x \dotplus 5/2^{n+1})}{2}\right| + \cdots + \left|\frac{f(x \dotplus (2^{n+1}-2)/2^{n+1}) - f(x \dotplus (2^{n+1}-1)/2^{n+1})}{2^n - 1}\right|.$$

Then if $U_n(x)$ converges uniformly to zero as n increases, $s_n(x, f)$ converges uniformly to f.

From this result we can easily derive the Dini-Lipschitz and Jordan tests for uniform convergence. An analogue of Theorem 3 on bounded Φ-variation also holds for Walsh-Fourier series as a consequence of Theorem 5.

The validity of the theorem of Salem for the characters of the group 2^ω suggests the problem of extending that result to other groups.

Let us suppose that G is a totally disconnected compact abelian group satisfying the second axiom of countability. Let X be its character group; then X is a discrete, countable, abelian torsion group. X can be written as the union of an increasing sequence of subgroups X_n such that $X_0=\{\chi_0\}$, where χ_0 is the identity character, and X_n/X_{n-1} is cyclic of prime order p_n. We suppose $\sup p_n = p<\infty$. The X_n are chosen so that there is a sequence $\{\phi_n\}$ of characters such that $\phi_n \in X_{n+1} \sim X_n$ and $\phi_n^{p_{n+1}} \in X_n$. Let $m_n=\prod_1^n p_i$, $m_0=1$. We may order X as follows. Each non-negative integer k has a unique representation, $k=\sum_{i=0}^s a_i m_i$ with $0 \leqq a_i < p_{i+1}$ for $0 \leqq i \leqq s$. We define

$$\chi_k = \phi_0^{a_0} \phi_1^{a_1} \cdots \phi_s^{a_s}.$$

We note that $\chi_{m_n}=\phi_n$ and $X_n=\{\chi_i : 0 \leqq i < m_n\}$.

Now let G_n be the annihilator of X_n, i.e., $G_n=\{x \in G : \chi_k(x)=1 \text{ for } 0 \leqq k < m_n\}$. Then $G=G_0 \supset G_1 \supset G_2 \supset \cdots$ and $\bigcap_0^\infty G_n=\{0\}$. It is easily seen that the G_n form a basis for the neighborhoods of 0 in G. Choose $x_n \in G_n \sim G_{n+1}$ such that $\chi_{m_n}(x_n)=\exp(2\pi i/p_{n+1})$. Each $x \in G$ has a unique representation $x=\sum_0^\infty b_i x_i$, $0 \leqq b_i < p_{i+1}$. This representation defines an ordering in G by means of the lexicographical ordering of the sequences $\{b_i\}$. It is clear that

$$G_n = \left\{x \in G : x = \sum_0^\infty b_i x_i \quad \text{with } b_0 = \cdots = b_{n-1} = 0\right\}.$$

Thus each coset of G_n has a unique representation $z+G_n$, where $z=\sum_{i=0}^{n-1} b_i x_i$, $0 \leqq b_i < p_{i+1}$. We denote these z, ordered lexicographically, by $\{z_\alpha^{(n)}\}$, $0 \leqq \alpha < m_n$. Let $\omega_k=\exp(2\pi i/p_{k+1})$.

The Fourier series of an integrable function f is given by

$$\sum_0^\infty c_n \chi_n(x), \text{ where } c_n = \int_G f(t)\bar{\chi}_n(t)\,dt.$$

Fourier series over the group G have been studied by Vilenkin [V].

The following generalization of the theorem of Salem has been shown by Onneweer and Waterman [OW]:

THEOREM 6. *Suppose f is continuous on G. Then the Fourier series of f converges uniformly if*

$$\sum_{\alpha=1}^{m_k-1} \frac{1}{\alpha} \left| \sum_{j=0}^{p_{k+1}-1} f(u - z_\alpha^{(k)} - j x_k)\,\omega_k^{j a_k} \right|$$

converges to zero uniformly in u and $a_k \in \{1, 2, \cdots, p_{k+1}-1\}$ *as* $k \to \infty$.

From this result one can readily obtain tests for uniform convergence and results on bounded Φ-variation analogous to those already obtained for trigonometric and Walsh series. Onneweer [O_2] has also succeeded in extending Theorem 4 to this setting.

6. A theorem of Goffman and Waterman. Theorem 3 yields a wide class of functions with the property that the Fourier series of $f \circ g$ converges uniformly for every homeomorphism g. This suggests the question as to whether or not one can characterize the functions with this property. In this regard, a characterization has been obtained by Goffman and Waterman [GW_1] of the continuous functions f, with $f(-\pi)=f(\pi)$, for which the Fourier series of $f \circ g$ converges everywhere (perhaps not uniformly) for every homeomorphism g. It is obtained by an elaboration of the method of Salem.

We define right and left systems of intervals at a point x. Let $\{k_n\}$ be a sequence of positive integers such that $\lim_n k_n = \infty$ and $\lim_n (k_n/n)=0$. For each n and $m=1, \cdots, k_n$, let I_{nm} be disjoint closed intervals such that $I_{n,m-1}$ is to the left of I_{nm}. If for some real x the collection

$$\mathcal{I} = \{I_{nm}: n = 1, 2, \cdots, m = 1, \cdots, k_n\}$$

has the property that, for each $\epsilon > 0$, there is an N such that $I_{nm} \subset (x, x+\epsilon)$ for all $n > N$, then $\mathcal{I}$ is called a *right system of intervals at* x. For each right system $\mathcal{I}$, consider the sequence

$$\alpha_n(\mathcal{I}) = \sum_{i=1}^{k_n} \frac{1}{m} f(I_{nm}),$$

where $f(I)=f(b)-f(a)$ for any interval $I=[a, b]$. Left systems, $\mathcal{I}$, and the numbers $\alpha_n(\mathcal{I})$ are defined similarly.

An elaboration and suitable modification of the technique of the Salem proof shows that if f is continuous, $f(-\pi)=f(\pi)$, and $\lim_n \alpha_n(\mathcal{I})=0$ for every right and left system $\mathcal{I}$, then the Fourier series of f converges everywhere. If, conversely, there is an $\mathcal{I}$ for which $\lim_n \alpha_n(\mathcal{I}) \neq 0$, then there is a homeomorphism g such that the Fourier series of $f \circ g$ diverges somewhere. Concerning the construction of g, we only note that the following lemma is used. The symbol $\sum(f, k, n, \theta)$ denotes

$$\sum_{i=1}^{k_n} \frac{1}{i}\left[f\left(\frac{2i\pi}{n}+\theta\right) - f\left(\frac{(2i-1)\pi}{n}+\theta\right)\right].$$

LEMMA 1. *Let* $\{k_n\}$ *be a sequence of integers with the properties* $\lim_n k_n = \infty$ *and* $\lim_n k_n/n = 0$. *There is a sequence* $\{\epsilon_n\}$, $0<\epsilon_n<\pi/n$, *such that for every function* h, *continuous in a neighborhood of zero, with* $h(0)=0$, *there is a sequence* $\{\theta_n\}$, $0<\theta_n<(\pi/n)-\epsilon_n$, *such that*

$$\int_0^{((2k_n+1)\pi)/n} h(t)\frac{\sin nt}{t}\,dt \text{ and } (1/\pi)\sum(h, k_n, n, \theta_n) \text{ are equiconvergent.}$$

We thus have the following result:

THEOREM 7. *A continuous f, with $f(-\pi)=f(\pi)$, is such that $f \circ g$ has an everywhere convergent Fourier series for each homeomorphism g if and only if $\lim_n \alpha_n(\mathcal{G})=0$ for every right and left system $\mathcal{G}$.*

The analogous question with uniform convergence remains open. There are continuous functions whose Fourier series converge everywhere without converging uniformly [**Zy**]. Is there a continuous function f such that the Fourier series of $f \circ g$ converges everywhere, for every homeomorphism g, but does not converge uniformly for some g?

7. The Menchov Theorem. We turn now to the second way in which the divergence of the Fourier series of a continuous function may be corrected. This is given by the following theorem of Menchov [**Ba**].

THEOREM 8. *If f is a measurable function on $[-\pi, \pi]$ and $\epsilon>0$, then there is a continuous function g such that $f(x)=g(x)$ except on a set of measure less than ϵ, and the Fourier series of g converges uniformly.*

The proof of this theorem is quite complicated. There seems to be little hope of proving it by finding a g which solves the problem by being in a class for which Fourier series converge uniformly. In particular, for each modulus of continuity ω, there is a continuous f such that any g whose modulus is ω differs from f almost everywhere, [**G**]. This theorem also holds for Walsh series, the proof being similar in principle and more transparent. We shall accordingly discuss this case, which was given by Kotlyar [**K**] and in sharper form by Price [**Pr**].

We indicate the argument given by Price. Let $V(n)$ be the vector space determined by the functions ψ_i, $i<2^n$, and for $n<k$, let $V(n, k)$ be the vector space determined by the functions ψ_i, $2^n \leqq i < 2^k$. Then $V(n, k)$ is the orthogonal complement of $V(n)$ in $V(k)$. Let g_{nj} be the characteristic function of the interval $[j/2^n, (j+1)/2^n]$. The following lemma allows enough freedom to prove the adjustment theorem:

LEMMA 2. *For each g_{nj}, each $N \geqq n$, and each $r>0$, there is a function $g \in V(N, N+r)$, which vanishes off $[j/2^n, (j+1)/2^n]$, is equal to g_{nj} except on a set of measure less than $1/2^{n+r}$, and is such that g and the partial sums of its Walsh series are all bounded by 2^r.*

By an ingenious iterated application of this lemma, using the fact that, for each n, the corresponding N can be taken to be arbitrarily large, Price obtains the following noteworthy result.

THEOREM 9. *If f is measurable and finite a.e. on $[0, 1]$, then for each $\epsilon>0$ and*

$$p_1 < q_1 < \cdots < p_i < q_i < \cdots,$$

where $\{q_n - p_n\}$ is increasing, there is a set E of measure less than ϵ, and a function g, which differs from f only on E, whose Walsh-Fourier series converges uniformly

and involves only Walsh functions ψ_n for which $p_i < n < q_i$. Moreover, for continuous f, the set E depends only on the modulus of continuity of f.

The scope of validity of the phenomenon of Theorems 8 and 9 remains to be determined. In particular, it would be interesting to find out whether it belongs to either the theory of real functions or to harmonic analysis. That is to say, is it concerned with orthonormal systems, or with group characters?

8. Walsh-like systems. Another interesting problem is to study the effect of changes of variable on orthogonal systems. Suppose $\{\phi_n\}$ is a system of functions on (0, 1) with $\phi_0 \equiv 1$. From this system we form $\{\psi_n\}$ just as the Walsh functions are obtained from the Rademacher functions. Thus for $n = 2^{k_1} + \cdots + 2^{k_s}$ with $0 \leqq k_1 < k_2 < \cdots < k_s$, we set

$$\psi_n = \phi_{k_1} \cdot \phi_{k_2} \cdot \cdots \cdot \phi_{k_s}.$$

If $\{\psi_n\}$ is an orthogonal system on (0, 1) it is called a *W-system*. We assume also that there is a K and an n_0 such that

$$(*) \qquad \int_0^1 |\psi_n|^2 = K \quad \text{for } n \geqq n_0.$$

The study of these systems has shown that they are essentially of two types depending on whether or not we assume

$$(**) \qquad |\phi_n(x)| \leqq 1 \quad \text{a.e.} \quad \text{for every } n.$$

Under this assumption, the results on W-systems largely parallel those for the Walsh system [**A**, 185–196]. Without (**) the behavior may be like that of the systems generated by the strongly lacunary trigonometric sequences $\{2^{1/2} \cos m_n x\}$ and $\{2^{1/2} \sin m_n x\}$, with $m_{n+1}/m_n \geqq 3$ [**A**, 190–191].

It has been shown by Waterman [**W**] that a W-system satisfying (*) and (**) behaves like the Walsh system because it is obtained from the Walsh system by a suitable change of variable.

It is not difficult to see that there is a set E of measure K in (0, 1) such that

(1) $|\psi_n| = 1$ on E for every n,

(2) $m(\{\psi_n \neq 0\} \cap E^c) > 0$ for only finitely many n,

(3) $\{\psi_n\}$ is orthogonal relative to E.

If we now let A be a 1-1 measure preserving map of (0, 1) onto itself such that $A((0, K)) = E$, we see that $\{\psi_n \circ A(Kx)\}$ is a W-system on (0, 1) and $|\psi_n \circ A(Kx)| = 1$ a.e. for all n. We may therefore assume $|\phi_n(x)| = 1$ a.e. for all n. We can show that we can define a function on (0, 1) by means of the dyadic representation of its values,

$$y(x) = \cdot \alpha_1 \alpha_2 \alpha_3 \cdots$$

with $\alpha_\nu = (1/\pi) \log \phi_\nu(x)$. It may be shown that for every measurable $E \subseteq (0, 1)$, $y^{-1}(E)$ is measurable and $m(y^{-1}(E)) = m(E)$. Clearly

$$\psi_n(x) = w_n \circ y(x) \quad \text{a.e. for all } n.$$

Further we have the following result:

THEOREM 10. *$\{\psi_n\}$ is complete if and only if there is a metric automorphism η of* $(0, 1)$ *such that* $\psi_n(x) = w_n \circ \eta(x)$ *a.e. for every n.*

Stein [**St**] has shown that there is a summable function whose Walsh-Fourier series diverges a.e., and Billard [**Bi**] has shown that the Walsh-Fourier series of a function of class L^2 converges a.e. As corollaries of these results and Theorem 10 we obtain:

COROLLARY 1. *For any W-system there is an integrable function whose W-Fourier series diverges a.e.*

COROLLARY 2. *The W-Fourier series of an L^2 function converges a.e.*

9. Almost everywhere convergence. Menchov proved Theorem 8 in connection with a problem of Lusin on convergence of trigonometric series. Lusin's question was the following: If f is a measurable function, is there a trigonometric series, with coefficients converging to zero, which converges almost everywhere to f? This was answered in the affirmative by Menchov [**Ba**] for functions which are finite almost everywhere, using Theorem 8 as a tool in the proof. An analogous fact has not been established for the Walsh functions, and seems to be difficult. Thus, the theorem of Section 7 seems to be easier for Walsh series than for trigonometric series, while the theorem on the existence of a series converging almost everywhere to a given measurable function seems to be harder for Walsh series.

Regarding the almost everywhere convergence of the Fourier series of a function in L^2, the proof has been recently given for Walsh series by Billard [**Bi**]. The proof, though by no means easy, is substantially less difficult than for the trigonometric case. Along the same lines, the theorem that a Fourier series is always (C, α) summable, $\alpha > 0$, turned out to be a major effort for the case of Walsh series. It was proved by Fine [$\mathbf{F_2}$].

The only complete orthonormal systems for which it is known that every finite measurable function is given by an almost everywhere convergent series are the trigonometric functions and the Haar functions. We recall that the *Haar functions* are the following complete orthonormal set on the interval $[0, 1]$:

$$h_{00}(x) \equiv 1, \qquad h_{01}(x) = \begin{cases} 1, & x \in [0, \frac{1}{2}) \\ -1, & x \in (\frac{1}{2}, 1] \\ 0, & x = \frac{1}{2} \end{cases},$$

$$h_{nk}(x) = \begin{cases} \sqrt{2^n}, & x \in \left[\dfrac{2k-2}{2^{n+1}}, \dfrac{2k-1}{2^{n+1}}\right) \\ -\sqrt{2^n}, & x \in \left(\dfrac{2k-1}{2^{n+1}}, \dfrac{2k}{2^{n+1}}\right] \\ 0, & \text{elsewhere}, \end{cases} \qquad \begin{matrix} n = 1, 2, \cdots \\ k = 1, \cdots, 2^n. \end{matrix}$$

The fact that, for each finite measurable f, there is a Haar series which converges almost everywhere was obtained by Bari [$\mathbf{T}_1$] as a consequence of the theorem of Lusin [**H**] which asserts that, for every such f, there is a continuous F such that $F'=f$ almost everywhere. The desired Haar series has as its coefficients rather evident linear combinations of the values of F at end points of dyadic intervals. By modifying F by a suitable Cantor function, it is clear that we may omit any finite set of Haar functions and obtain, for each finite measurable f, a Haar series which converges almost everywhere to f and involves only the remaining Haar functions. Whether or not certain infinite sets of Haar functions may be omitted has not been considered.

The fact that a finite measurable function may be represented as an almost everywhere limit of Haar functions, with a finite set of functions omitted, assures the existence of sets of multiplicity of measure 0 for the Haar functions. We recall that a set S is a *set of uniqueness* for a system $\{\phi_n\}$ if the convergence to zero of $\sum_{n=1}^{\infty} a_n\phi_n$ on the complement of S implies that the coefficients a_n are all zero. Also S is a *set of multiplicity* if it is not a set of uniqueness. Since the above implies the existence of a nonzero Haar series converging almost everywhere to zero, it assures the existence of sets of multiplicity of measure zero.

Whether or not every measurable function is the almost everywhere limit of a series of trigonometric functions with finitely many omitted, is not known.

10. Convergence in measure. The original Lusin problem included functions which may be infinite on sets of positive measure. It has recently been shown by Arutyanyan and Talalyan [**AT**] that if f is infinite on a set of positive measure, then no Haar series converges almost everywhere to f; and the same applies to Walsh series. Considerable light has been cast on this matter by Gundy [**Gu**], who has related it to martingale theory.

In regard to Lusin's question, Menchov asked whether the answer is affirmative using convergence in measure instead of almost everywhere convergence. He showed [**M**] that for each measurable function f, finite or infinite, there is a trigonometric series which converges in measure to f. The proof given by Menchov is grimly complicated. Fortunately, Talalyan [$\mathbf{T}_1$] gave an elegant, understandable, although by no means easy, proof of the theorem in a more general setting.

THEOREM 11. *If $\{\phi_n\}$ is a sequence in $L^p[a, b]$, $p>1$, of functions of norm* 1, *which is a Schauder basis for $L^p[a, b]$ then, for each measurable f, finite or infinite, there is a series $\sum_{n=1}^{\infty} a_n\phi_n$, with* $\lim_n a_n=0$, *which converges in measure to f.*

A central role in the proof of this theorem is assumed by a lemma.

LEMMA 3. *If G is a bounded measurable set of positive measure, $\{\phi_n\}$ is a sequence in $L^p(G)$, $p>1$, of functions of norm* 1, *which is a Schauder basis for $L^p(G)$, and f is finite and measurable, then for each $\epsilon>0$ and positive n, there is a measurable set $E\subset G$ with $m(E)<\epsilon$, and there are reals a_k $(k=n+1, \cdots, m)$ with*

$|a_k| < \epsilon$ *such that*

$$\left\| \sum_{k=n+1}^{m} a_k\phi_k - f \right\|_{G-E} < \epsilon \quad \textit{and} \quad \left\| \sum_{k=n+1}^{s} a_k\phi_k \right\|_{F} < \|f\|_F + \epsilon,$$

for each measurable $F \subset G-E$ *and* $s = n+1, \cdots, m$.

From the manner in which this lemma is used in the proof of the theorem, it follows that any finite subset of $\{\phi_n\}$ can be dropped and the theorem remains valid. It is again not clear whether or not certain infinite sets can be dropped.

It seems that the phenomenon of Theorem 11 should have no relation to Schauder bases in L_p. However, the proof of Talalyan is rigidly bound to this special class of systems. A new method would be needed to go further.

The Lemma 3 has further interest. Indeed, the existence of universal series for all possible systems was established by Talalyan [**T**$_2$] with its use. It turns out, however, that the existence of such series is quite easy to establish, as noted by Goffman and Waterman [**GW**$_2$]. We discuss the matter in Section 12.

11. The Müntz Theorem. We consider first the functions $1, x, \cdots, x^n, \cdots$ defined on $[0, 1]$. The Weierstrass polynomial approximation theorem asserts that each continuous function on $[0, 1]$ is the uniform limit of finite linear combinations of these functions. On the other hand, only analytic functions are uniform limits of series $\sum_{n=0}^{\infty} a_n x^n$. A similar fact holds for $L^p[0, 1]$, $p \geqq 1$. Indeed, if $\sum_{n=0}^{\infty} a_n x^n$ converges at $x \neq 0$, it must converge uniformly to an analytic function on each interval $(-|x|+\delta, |x|-\delta)$ for small $\delta > 0$.

Some of the functions in the set $1, x, x^2, \cdots$ may be removed, and each continuous function is still a uniform limit of finite linear combinations of the remaining functions. We have the marvelous theorem of Müntz [**Fe**]:

The finite linear combinations of $1, x^{n_1}, \cdots, x^{n_k}, \cdots$ *are dense in* $C[0, 1]$ *if and only if* $\sum_{k=1}^{\infty} 1/n_k = \infty$.

We say that a sequence is *total* in a space X if its finite linear combinations are dense in X. Müntz's Theorem holds in $L^2[0, 1]$, with the difference that the function 1 does not play a special role. Thus $x^{n_1}, \cdots, x^{n_k}, \cdots$ *is total in* $L^2[0, 1]$ *if and only if* $\sum_{k=1}^{\infty} 1/n_k = \infty$. The same holds for any $L^p[0, 1]$, $p \geqq 1$.

If $\sum_{k=1}^{\infty} 1/n_k < \infty$, the uniform closure of span of $x^{n_1}, \cdots, x^{n_k}, \cdots$ is a subset of the set of analytic functions. This fact was proved by Clarkson and Erdös [**CE**]. In the same work, they showed that Müntz's Theorem holds for any positive interval in the place of $[0, 1]$. It would be interesting to know if it holds for arbitrary measurable subsets of $[0, 1]$ of positive measure. This would imply that the theorem holds for the space M of equivalence classes of measurable functions, with almost everywhere convergence or convergence in measure.

Müntz Theorem assures the existence of a universal series. That is to say, there is a series

$$a_0 + a_1 x + \cdots + a_n x^n + \cdots$$

such that, if $\{s_n\}$ is its sequence of partial sums, each $f \in L^2[0, 1]$ is the limit, in the L_2 norm, of a subsequence of $\{s_n\}$. Universal series exist in the space of measurable functions for every total set.

12. The space *M*. We consider the space M of measurable functions on $[0, 1]$, with the convergence in measure metric. This is a topological vector space which is not locally convex. It has a dual, the space of continuous linear functionals, which consists of the zero functional alone. In order to see this, let ϕ be a nonzero linear functional on M. There is then a positive f such that $\phi(f) \neq 0$. For each n, and $i = 0, \cdots, n-1$, let s_{ni} be the characteristic function of $[i/n, (i+1)/n]$. There is an $i = i_n$ and a real number a_n such that $\phi(a_n s_{ni_n} f) = 1$. Now $\{a_n s_{ni_n} f\}$ converges to zero in M, so that ϕ is not continuous.

We also need the fact that if $\{x_n\}$ is total in a topological vector space X, and Y is the closed span of $\{x_2, \cdots, x_n, \cdots\}$ in X, then Y has deficiency 0 or 1.

From these two facts, we almost immediately have the following result:

THEOREM 12. *If S is total in M, and F is a finite subset of S, then $S - F$ is total in M.*

The two corollaries below follow readily.

·COROLLARY 1. *If S is total in M, there is an infinite $T \subset S$ such that $S - T$ is total in M.*

COROLLARY 2. *If $\{\phi_n\}$ is total in M, there is a series $\sum_{n=1}^{\infty} a_n \phi_n$ whose sequence of partial sums has the property that, for each measurable f, a subsequence converges almost everywhere to f.*

Little is known concerning the related matter for series. We pointed out that if a finite set of Haar functions is removed, every $f \in M$ has a series in the remaining Haar functions which converges almost everywhere to f. No analogous property is known for trigonometric functions. The general problem has connections with the existence of sets of multiplicity, as has already been noted.

For the case of the Schauder functions, a system which is total in each $L_p[0, 1]$, $p > 1$, but which will not be defined here, Zink [Z] has given necessary and sufficient conditions that a subsequence $\{\phi_{n_k}\}$ of the Schauder system $\{\phi_n\}$ be such that every measurable function f, finite or infinite, have a series expansion

$$\sum a_k \phi_{n_k}$$

which converges almost everywhere to f.

References

[A] G. Alexits, Convergence Problems of Orthogonal Series, Hungarian Acad. of Sci., Budapest, 1961.

[AT] F. G. Arutyanyan and A. A. Talalyan, On convergence of Haar series to $+\infty$, (Russian) Mat. Sbornik, 66 (1965) 240–247.

[Ba] N. K. Bary, A Treatise on Trigonometric Series, Macmillan, New York, 1964.

[Bi] P. Billard, Sur la convergence presque partout des séries de Fourier-Walsh des fonctions de l'espace $L^2(0, 1)$, Studia Math, 28 (1966) 363–388.

[B] H. Bohr, Über einen Satz von J. Pál, Acta Szeged., 7 (1935) 129–135.

[CE] J. A. Clarkson and P. Erdős, Approximation by polynomials, Duke Math. J., 10 (1943) 5–11.

[Fe] W. Feller, On Müntz theorem and completely monotone functions, this MONTHLY, 75 (1968) 342–350.

[F_1] N. Fine, On the Walsh functions, Trans. A.M.S., 65 (1949) 372–414.

[F_2] N. Fine, Cesàro summability of Walsh-Fourier series, Proc. Nat. Acad. USA, 41 (1955) 588–591.

[FR] W. H. Fleming and R. Rishel, An integral formula for total gradient variation, Arch. Math., 11 (1960) 218–222.

[G] C. Goffman, Approximation of non-parametric surfaces of finite area, J. Math. Mech., 12 (1963) 737–745.

[GS] A. M. Garsia and S. Sawyer, On some classes of continuous functions with convergent Fourier series, J. Math. Mech., 13 (1964) 589–601.

[Gu] R. F. Gundy, Martingale theory and pointwise convergence of certain orthogonal series, Trans. A.M.S., 124 (1966) 228–248.

[GW_1] C. Goffman and D. Waterman, Functions whose Fourier series converge for every change of variable, Proc. A.M.S., 19 (1968) 80–86.

[GW_2] C. Goffman and D. Waterman, Basic sequences in the space of measurable functions, Proc. A.M.S., 11 (1960) 211–213.

[H] E. Hobson, The theory of Functions of a Real Variable, Dover, New York, 1957.

[K] B. D. Kotljar, Walsh series and the theorem of Menchov on "correction" of functions, Izv. Akad. Nauk, SSSR, 30 (1966) 1193–1200.

[L] H. Lebesgue, Sur les intégrales singulières, Ann. Fac. Sc. Univ. Toulouse (3), I (1909) 25–117.

[M] D. E. Menchov, On the convergence in measure of trigonometric series, Trudy Matem. Inst. i.V.A. Steklova, 1950 (A.M.S. Translation No. 105).

[O_1] C. W. Onneweer, On uniform convergence for Walsh-Fourier series, submitted to Pac. J.

[O_2] C. W. Onneweer, Ph.D. dissertation, Wayne State U., 1969.

[OW] C. W. Onneweer and D. Waterman, unpublished manuscript.

[P] J. Pàl, Sur des transformations de fonctions qui font converger leurs séries de Fourier, C.R., 158 (1914) 101–103.

[Pr] J. J. Price, Walsh series and adjustment of functions on small sets, Ill. J. Math., 13 (1969) 131–136.

[S] R. Salem, Essais sur les séries trigonométriques, Act. Sc. et Ind., no. 862, Paris, 1940.

[St] E. M. Stein, On limits of sequences of operators, Ann. of Math., 74 (1961) 140–170.

[T_1] A. A. Talalyan, Representation of measurable functions by series, Russian Math. Surveys, 15 (1960) 75–136.

[T_2] A. A. Talalyan, On the convergence almost everywhere of subsequences of partial sums of general orthogonal series, (Russian) Akad. Nauk Arm. SSR, 10 (1957), 17–34.

[V] N. Ja. Vilenkin, On a class of complete orthonormal systems, Izv. Akad. Nauk SSSR Ser. Mat., 11 (1947) 363–400 (A.M.S. Translations Ser. 2, vol. 28, 1–35).

[W] D. Waterman, W-systems are the Walsh functions, Bull. A.M.S. 75 (1969) 139–142.

[Y] L. C. Young, Sur une généralisation de la notion de variation de puissance p-ième bornée au sens de M. Wiener, et sur la convergence des séries de Fourier, C.R., 204 (1937) 470–472.

[Z] R. Zink, On a theorem of Goffman concerning Schauder series, Proc. A.M.S., 21 (1969) 523–529.

[Zy] A. Zygmund, Trigonometric Series, Cambridge Univ. Press, Cambridge, 1959.

THE WORK OF NICHOLAS BOURBAKI*

JEAN A. DIEUDONNÉ (translated by Linda Bennison)

Dear President and friend,
Ladies and Gentlemen,

Thank you for your kind words on my behalf. I must admit that it is a great pleasure for me to escape from my duties as dean and spend a week in such a warm and friendly atmosphere. The tradition of friendship between French and Roumanian scientists, particularly between our mathematicians, is old. I am very happy to be a link in this chain, which I hope will continue, stronger and more cordial, in the years to come. Well, if you don't mind, I shall not make a very long speech, and I should be happy to answer at the end of it the questions which, no doubt, will be put to me. I don't pretend to deal with all the history of the works of Bourbaki, and I shall give you all a chance to ask questions on the points I merely touch upon.

To understand the origins of Bourbaki, we shall have to go back to the years that Mr. Nicolescu was recalling a few moments ago. These were the years when we were students, the years after the 1914 war; and this war, we can very well say, was extremely tragic for the French mathematicians. I shall not try to judge or give a moral assessment of what happened at that time. In the great conflict of 1914–18, the German and French governments did not see things in the same way where science was concerned. The Germans put their scholars to scientific work, to raise the potential of the army by their discoveries and by the improvement of inventions or processes, which in turn served to augment the German fighting power. The French, at least at the beginning of the war and for a year or two, felt that everybody should go to the

---

It is hardly necessary to identify Prof. Dieudonné to our readers; still a few facts may prove interesting. Prof. Dieudonné studied at the Ecole Normale supérieure from 1924–1927, was a fellow at Princeton, Berlin, and Zürich, and received his doctorate in 1931. He served on the faculties at Bordeaux, Rennes, Nancy, Sao-Paulo, Michigan, Northwestern, *l'Institut des Hautes Etudes Scientifiques*, and presently is the Dean of the Faculty at Nice. He held visiting professorships at Columbia, Johns Hopkins, Rio de Janeiro, Buenos Aires, Pisa, Maryland, Tata Institute Bombay, Notre Dame, and Washington. His honors include the Order of the Legion of Honor, the Order of the Academic Palms, and membership in the Academy of Sciences. He served as President of the Mathematical Society of France in 1964–65.

Prof. Dieudonné has published a number of books and about 135 research articles on analysis, topology, spectral theory, classical groups, formal Lie groups, and non-commutative rings.

It is believed by some that the original Bourbaki members were C. Chevalley, J. Delsarte, J. Dieudonné, and A. Weil. Apparently the name "Nicolaus Bourbaki" was that of a 19th century French general. It was part of an "initiation" of first year mathematics students that an upper classman, pretending to be a famous foreign mathematician, would deliver a lecture to them in which the theorems bore the names of famous generals and were all wrong in a nontrivial way. *Editor.*

front; so the young scientists, like the rest of the French, did their duty at the front line. This showed a spirit of democracy and patriotism that we can only respect, but the result was a dreadful hecatomb of young French scientists. When we open the war-time directory of the Ecole Normale, we find enormous gaps which signify that two-thirds of the ranks were mowed down by the war. This situation had unfortunate repercussions for French mathematics. We others, too young to have been in direct contact with the war, but entering the University in the years after the war ended, should have had as our guides these young mathematicians, certain of whom we are sure would have had great futures. These were the young men who were brutally decimated and whose influence was destroyed.

Obviously, people of previous generations were left, great scholars whom we all honour and respect. Masters like Picard, Montel, Borel, Hadamard, Denjoy, Lebesgue, etc., were living and still extremely active, but these mathematicians were nearly fifty years old, if not older. There was a generation between them and us. I am not saying that they did not teach us excellent mathematics: we all took first-class courses from these mathematicians (as Mr. Nicolescu is a witness), but it is indubitable (and true for that matter of every period) that a 50-year-old mathematician knows the mathematics he learned at 20 or 30, but has only notions, often rather vague, of the mathematics of his epoch, i.e., the period of time when he is 50. It is a fact we have to accept such as it is, we cannot do anything about it.

So we had excellent professors to teach us the mathematics of let us say up to 1900, but we did not know very much about the mathematics of 1920. As I said before, the Germans went about things in a different way, so that the German mathematics school in the years following the war had a brilliance which was altogether exceptional. We only need to think of the mathematicians of the highest order who illustrated this point: C. L. Siegel, E. Noether, E. Artin, W. Krull, H. Hasse, etc., of whom we in France knew nothing. Not only this, but we also knew nothing of the rapidly developing Russian school, the brilliant Polish school, which had just been born, and many others. We knew neither the work of F. Riesz nor that of von Neumann, etc. We had been closed in on ourselves and, in our world, the theory of functions reigned supreme. The only exception was Elie Cartan; but being 20 years ahead of his time, he was understood by no one. (The first to understand him after Poincaré was Hermann Weyl, and for 10 years he was the only one, so how could we poor little students have known enough to understand him?) So, apart from E. Cartan, who at this time didn't count—he only started to count 20 years later, but since then his influence has grown steadily—we were entirely folded in on that theory of functions, which, while being important, represented only a part of mathematics.

Our only opening onto the outside world at this time was the seminar of Hadamard, a professor, but not a very brilliant teacher, at the Collège de France. (He was a great enough scholar for me to be able to say this without

harming his reputation.) He had the idea (apparently taken from abroad, because this had never been done in France) of inaugurating a seminar of analysis of current mathematical work. At the beginning of the year he distributed, to all those who wanted to speak on the subject, what he judged to be the most important memoirs of the past year, and they had to explain them at the blackboard. It was a novelty for the time, and to us an extremely precious one, because there we met mathematicians of many different origins. Also, it soon became a center of attraction for foreigners; they came in crowds. (Mr. Onicescu reminded me that he himself gave lectures at the Hadamard seminar in Paris.) So it was for us young students a source of acquaintances and views that we did not find in the formal mathematics courses given at the University. This state of affairs lasted several years, until certain of us—starting with A. Weil, then C. Chevalley, having been out of France meeting Italians, Germans, Poles, etc.—realized that if we continued in this direction, France was sure to arrive at a dead end. We would no doubt continue to be very brilliant in the theory of functions, but for the rest, French mathematicians would be forgotten. This would break a two-hundred-year-old tradition in France, because from Fermat to Poincaré, the greatest of the French mathematicians had always had the reputation of being universal mathematicians, as capable in arithmetic as in algebra, or in analysis, or in geometry. So we had this warning of the bubbling of ideas that was beginning to be seen outside, and several of us had the chance to go and see and learn at first hand the development that was going on outside our walls. After Hadamard retired in 1934, the seminar was carried on, in a slightly different form, by G. Julia. This consisted of studying in a more systematic manner the great new ideas which were coming in from all directions. This is when the idea of drawing up an overall work which, no longer in the shape of a seminar, but in book form, would encompass the principal ideas of modern mathematics. From this was born the Bourbaki treatise. I must say that the collaborators of Bourbaki were very young at the time and doubtless they would never have started this job had they been older and better informed. In the first meetings for the project, the idea was that it would be finished in three years, and in this time we should draft the basic essentials of mathematics. Events and history decided differently. Little by little, as we became rather more competent and more aware, we realized the enormity of the job that had been taken on, and that there was no hope of finishing it as quickly as that.

It is true that there were already excellent monographs at the time and, in fact, the Bourbaki treatise was modelled in the beginning on the excellent algebra treatise of Van der Waerden. I have no wish to detract from his merit, but as you know, he himself says in his preface that really his treatise had several authors, including E. Noether and E. Artin, so that it was a bit of an early Bourbaki. This treatise made a great impression. I remember it—I was working on my thesis at that time; it was 1930 and I was in Berlin. I still remember the day that Van der Waerden came out on sale. My ignorance in algebra was such

that nowadays I would be refused admittance to a university. I rushed to those volumes and was stupefied to see the new world which opened before me. At that time my knowledge of algebra went no further than *mathématiques spéciales*, determinants, and a little on the solvability of equations and unicursal curves. I had graduated from the Ecole Normale and I did not know what an ideal was, and only just knew what a group was! This gives you an idea of what a young French mathematician knew in 1930. So we tried to follow Van der Waerden, but in effect he only covered algebra, and even then just a small part of algebra. (Since then, algebra has developed considerably, partly because of Van der Waerden's treatise, which is still an excellent introduction. I am often asked for advice on how to start out studying algebra, and to most people I say: First read Van der Waerden, in spite of what has been done since.)

So we intended to do something of this kind. Now Van der Waerden uses very precise language and has an extremely tight organization of the development of ideas and of the different parts of the work as a whole. As this seemed to us to be the best way of setting out the book, we had to draft many things which had never before been dealt with in detail. General topology could only be found in a few memoirs and in Fréchet's book, which was, in effect, a compilation of an enormous quantity of results, without any kind of order. I can say the same of Banach's book, which is admirable for research but completely disorganized; in other subjects such as integration (as presented by Bourbaki) and certain algebra questions, there was nothing. Before the chapter of Bourbaki on multilinear algebra, I don't think there was a didactic work in the world that explained what exterior algebra was. We had to refer to the work of Grassmann, which is not particularly clear. Thus we quickly realized that we had rushed into an enterprise which was considerably more vast than we had imagined, and you know that this enterprise is still far from finished. In my briefcase I have the proofs of the 34th volume, which is devoted to three chapters of the theory of Lie groups. There are others, many others, being prepared; there are already three or four editions of preceding volumes, and the end of the work is not in sight.

We had to have a starting point—we had to know what we wanted to do. Of course, there was the idea of the Encyclopedia, which, in fact, already existed. As you know, it had been started by the Germans in 1900, and despite their proverbial tenacity and ardour for work, in 1930, after several editions and alterations, etc., it was hopelessly behind in comparison to the mathematical science of that time. Nowadays, nobody would think of starting on such an impossible enterprise, knowing the vast amount of mathematical publications released every year. I believe that we shall have to wait for the day when computers have minds and are able to assimilate all that in a few minutes. For the time being we have not progressed that far, nor had we gone that far in 1930. Moreover, it would have been useless to redo something which despite its merits had failed. The Encyclopedia, even at that period, was above all useful as a

bibliographical reference, to find out where such and such a result could be found. But naturally, it contained no proofs, because if the Encyclopedia, already gigantic with its 25–30 volumes, had included proofs it would have been ten times larger. No, we did not want to produce a work of bibliographic reference, but one which would be a demonstrative mathematical text from beginning to end. And this forced us into making an extremely strict selection. What selection? Well, that is the crucial part in Bourbaki's evolution. The idea which soon became dominant is that the work had to be primarily a *tool*. It had to be something usable not only in a small part of mathematics, but also in the greatest possible number of mathematical places. So if you like, it had to concentrate on basic mathematical ideas and essential research. It had to reject completely anything secondary that had no immediately known application and that did not lead directly to conceptions of known and proved importance. There was much sifting, which started innumerable discussions among the collaborators, and which also earned Bourbaki a great deal of hostility. Because as the works of Bourbaki became known, all those who found that their favourite subject was not included were not inclined to do much propaganda in his favor. So I think that we can attribute much of the hostility that has been shown toward Bourbaki at certain periods, and which is still widespread in certain countries, to this extremely strict selection.

So how do we choose these fundamental theorems? Well, this is where a new idea came in: that of *mathematical structure*. I do not say it was an original idea of Bourbaki—there is no question of Bourbaki's containing anything original. Bourbaki does not attempt to innovate mathematics, and if a theorem is in Bourbaki, it was proved 2, 20, or 200 years ago. What Bourbaki has done is to define and generalize an idea which already was widespread for a long time. Since Hilbert and Dedekind, we have known very well that large parts of mathematics can develop logically and fruitfully from a small number of well-chosen axioms. That is to say, given the bases of a theory in an axiomatic form, we can develop the whole theory in a more comprehensible way than we could otherwise. This is what gave the general idea of the notion of mathematical structure. Let us say immediately that this notion has since been superseded by that of category and functor, which includes it under a more general and convenient form. It is certain that it will be the duty of Bourbaki, who, as I shall explain later, never fears change, to incorporate the valid ideas of this theory in his works.

Once this idea had been clarified, we had to decide which were the most important mathematical structures. Naturally, this was the root of many discussions before we found ourselves in agreement. I might say that Bourbaki does not pretend to be infallible; he has been mistaken several times about the future of structures, and apologized when it was necessary, withdrawing his original ideas. Successive editions trace some changes clearly. Bourbaki does not pretend to want to fix or nail down mathematics; that would be exactly contrary

to his original purpose. But if one does not recoil from new ideas, even when they go beyond Bourbaki, one has no respect for tradition. Consequently this open systematic attitude of Bourbaki has also been a cause of hostility, this time on the part of people of previous generations, who criticized the liberties Bourbaki took with the mathematics of their time. In particular, the choice of definitions and the order in which the subjects were arranged were decided according to a logical and rational scheme. If this did not agree with what was done previously, well, it means that what was done previously had to be thrown overboard, without sparing even long-established traditions. To give you an example: Bourbaki refuses to say *non-decreasing* when referring to an increasing function because this would be a total absurdity. We know that this term means what we want to say only when talking about linear (total) order relations. (If one says non-decreasing in the setting of a non-linear order relation, this hardly means *increasing but not strictly increasing.*) So Bourbaki purely and simply abolished this terminology, as he did many others. He also invented terminology, using Greek when it was necessary, but also using many words from ordinary speech, which made traditionalists wince. They did not admit easily that what we now call *boule* or *pavé* used to be called *hypersphéroide* or *parallélotope*, and their reaction was: "This work is not to be taken seriously." A little book came out recently, which we liked very much. It is called "*Le Jargon des Sciences*" by Etiemble, vigilant guardian of the French language. He insists on preserving it in its original purity and is up in arms against the gibberish of most scientists. Happily, he makes an exception of French mathematicians, saying that they had the good sense to take simple, authentic French words from ordinary speech, sometimes changing their meaning. He cites attractive examples, recent titles such as *Platitude et privilège* and *Sur les variétés riemanniennes non suffisamment pincées*. This is the style in which Bourbaki is written—in a recognizable language and not in a jargon sprinkled with abbreviations, as in Anglo-Saxon texts where you are told about the C.F.T.C. which is related to an A.L.V. unless it is a B.S.F. or a Z.D., etc. After ten pages of this you have no idea what they are talking about. We think that ink is cheap enough to write things in full, with a well-chosen vocabulary.

I told you then that we made a selection. I shall explain this choice in more detail, using a metaphor. We realized very quickly that despite introducing the idea of structure, which was meant to clarify and separate things, mathematics refused to separate into small pieces. On the other hand, it was clear that the old divisions, Algebra, Arithmetic, Geometry, Analysis were out of date. We had no respect for them and abandoned them from the start, to the fury of many. For example, it is well known that euclidean geometry is a special case of the theory of hermitian operators in Hilbert spaces. The same goes for the theories of algebraic curves and numbers, which come essentially from the same structures. I compare the old mathematical divisions with the divisions of the ancient zoologists, who, seeing that a dolphin and a shark or a tuna-fish were

similar animals, said: These are fish because they all live in the sea and have similar shapes. It was quite a while before they realized that the structures of these animals were not at all similar, and they had to be classified very differently. Algebra, Arithmetic, Geometry and all that nonsense compare easily to this. One has to look at the structure of each theory and classify it in this way. In spite of everything though, it does not take long to make one realize that despite this effort towards the isolation of structures, they have a way of mixing very quickly and extremely fruitfully. One could say that the great ideas in mathematics have come when several very different structures met. So here is my picture of mathematics now. It is a ball of wool, a tangled hank where all mathematics react one upon another in an almost unpredictable way. Unpredictable, because a year almost never passes without our finding new reactions of this kind. And then, in this ball of wool, there are a certain number of threads, coming out in all directions and not connecting up with anything else. Well, the Bourbaki method is very simple—we cut the threads. What does this mean? Let us look at what remains; then we make a list of what remains and a list of what is eliminated. What remains: The archiclassic structures (I don't speak of sets, of course), linear and multilinear algebra, a little general topology (the least possible), a little topological vector spaces (as little as possible), homological algebra, commutative algebra, non-commutative algebra, Lie groups, integration, differentiable manifolds, riemannian geometry, differential topology, harmonic analysis and its prolongations, ordinary and partial differential equations, group representation in general, and in its widest sense, analytical geometry. (Here of course I mean in the sense of Serre, the only tolerable sense. It is absolutely intolerable to use *analytical geometry* for linear algebra with coordinates, still called analytical geometry in the elementary books. Analytical geometry in this sense has never existed. There are only people who do linear algebra badly, by taking coordinates and this they call analytical geometry. Out with them! Everyone knows that analytical geometry is the theory of analytical spaces, one of the deepest and most difficult theories of all mathematics.) Algebraic geometry, its twin sister, is also included, and finally the theory of algebraic numbers.

This makes an imposing list. Let us now see what is excluded. The theory of ordinals and cardinals, universal algebra (you know very well what that is), lattices, non-associative algebra, most general topology, most of topological vector spaces, most of the group theory (finite groups), most of number theory (analytical number theory, among others). The processes of summation and everything that can be called hard analysis—trigonometrical series, interpolation, series of polynomials, etc.; there are many things here; and finally, of course, all applied mathematics.

There I wish to explain myself a little. I absolutely do not mean that in making this distinction Bourbaki makes the slightest evaluation on the ingeniousness and strength of theories catalogued in this way. I am convinced that the theory of finite groups, for example, is at the present time one of the deepest

and richest in extraordinary results, while theories like non-commutative algebra are of medium difficulty. And if I had to make an evaluation I should probably say that the most ingenious mathematics is excluded from Bourbaki, the results most admired because they display the ingenuity and penetration of its discoverer.

We are not talking about classification then, the good on my right, the bad on my left—we are not playing God. I just mean that if we want to be able to give an account of modern mathematics which satisfies this idea of establishing a center from which all the rest unfolds, it is necessary to eliminate many things. In group theory, despite the extraordinary penetrating theorems which have been proved, one cannot say that we have a general method of attack. We have several of them, and one always has the impression that one is working like a craftsman, by accumulating a series of stratagems. This is not something which can be set forth by Bourbaki. Bourbaki can only and only wants to set forth theories which are rationally organized, where the methods follow naturally from the premises, and where there is hardly any room for ingenious stratagems.

So, I repeat, those which Bourbaki proposes to set forth are generally mathematical theories almost completely worn out already, at least in their foundations. This is only a question of foundations, not details. These theories have arrived at the point where they can be outlined in an entirely rational way. It is certain that group theory (and still more analytical number theory) is just a succession of contrivances, each one more extraordinary than the last, and thus extremely anti-Bourbaki. I repeat, this absolutely does not mean that it is to be looked down upon. On the contrary, a mathematician's work is shown in what he is capable of inventing, even new stratagems. You know the old story—the first time it is a stratagem, the third time a method. Well, I believe that greater merit comes to the man who invents the stratagem for the first time than to the man who realizes after three or four times that he can make a method from it. The second step is Bourbaki's aim: to gather from the diverse processes used by mathematicians whatever can be shaped into a coherent theory, logically arranged, easily set forth and easily used.

The work method used in Bourbaki is a terribly long and painful one, but is almost imposed by the project itself. In our meetings, held two or three times a year, once we have more or less agreed on the necessity of doing a book or chapter on such and such a subject (generally, we foresee a certain number of chapters for a book), the job of drafting it is put into the hands of the collaborator who wants to do it. So he writes one version of the proposed chapter or chapters from a rather vague plan. Here, generally, he is free to insert or neglect what he will, completely at his own risk and peril, as you will see. After one or two years, when the work is done, it is brought before the Bourbaki Congress, where it is read aloud, not missing a single page. Each proof is examined, point by point, and is criticized pitilessly. One has to see a Bourbaki

Congress to realize the virulence of this criticism and how it surpasses by far any outside attack. The language cannot be repeated here. The question of age does not come into it. The ages of the Bourbaki members vary considerably—later I shall tell you the maximum age limit—but even when two men have a 20-year age difference, this does not stop the younger from hauling the elder, who he feels has understood nothing of the question, over the coals. One has to know how to take it, as one should, with a smile. In any case, the reply is never late in coming, no one can boast of being infallible before Bourbaki members, and in the end, everything works out fine, despite the very long and extremely animated arguments.

Certain foreigners, invited as spectators to Bourbaki meetings, always come out with the impression that it is a gathering of madmen. They could not imagine how these people, shouting—some times three or four at the same time—about mathematics, could ever come up with something intelligent. It is perhaps a mystery but everything calms down in the end. Once the first version has been torn to pieces—reduced to nothing—we pick a second collaborator to start it all over again. This poor man knows what will happen because although he sets off following the new instructions, meanwhile the ideas of the Congress will change and next year *his* version will be torn to bits. A third man will start, and so it will go on. One would think it was an unending process, a continuous recurrence, but in fact, we stop for purely human reasons. When we have seen the same chapter come back six, seven, eight, or ten times, everybody is so sick of it that there is a unanimous vote to send it to press. This does not mean that it is perfect, and very often we realize that we were wrong, in spite of all the preliminary precautions, to start out on such and such a course. So we come up with different ideas in successive editions. But certainly the greatest difficulty is in the delivery of the first edition.

An average of 8–12 years is necessary from the first moment we set to work on a chapter to the moment it appears in the bookshop. The ones that are coming out now are the ones that were discussed for the first time about 1955.

I said earlier that there is a maximum age limit. This was recognized quite quickly for the reason I was speaking about at the start of this talk—a man of over 50 can still be a very good and extremely productive mathematician but it is rare for him to adapt to the new ideas, to the ideas of people 25 and 30 years younger than he. Now, an enterprise like Bourbaki seeks to be permanent. There is no question of saying that we nail down mathematics to such or such a period. If the mathematics set forth by Bourbaki no longer corresponds to the trends of the period, the work is useless and has to be redone. This has already happened, for that matter, with several volumes of Bourbaki. If there were elderly members of Bourbaki, they would tend to put a brake on this healthy tendency, believing that everything being fine at the time of their youth, there is no reason for change. This would be disastrous. So, to avoid tensions such as this, which sooner or later would cause Bourbaki's break-up, it was decided at the time

the question arose, that all the Bourbaki collaborators retire at 50.

And it is so; the present Bourbaki collaborators are all under 50. The founder-members, of course, retired almost ten years ago, and even those who not long ago were considered young are already past—or about to reach—retiring age. So it is a question of replacing the members who leave. How do we do that? Well, there are no rules, because in Bourbaki the only formal rule is the one I have just told you, retirement at 50. Apart from this, we can say that the only rule is that there are no rules. There are no rules in the sense that there is never a vote, we have to have unanimity on every point. Each member has the right to veto any chapter he feels is bad. The veto simply signifies that we do not allow the printing of the chapter and we have to go back and re-study it. This explains the lengthiness of the process—the fact that we have such a hard time agreeing on a final version.

We are concerned then with replacing members affected by the age limit. We do not replace them formally (this would be a rule and there are no rules). There is no vacant seat, as with an academy. As most of the members of Bourbaki are professors—many in Paris—they have a chance to see at close range the young mathematicians, the youths who are just starting mathematical research. A youth of value who shows promise of a great future is quickly noticed. When this happens, he is invited to attend one of the Congresses as a guinea pig. This is the traditional method. You all know what a guinea pig is—the small animal that we use to test all the viruses. Well, it is much the same thing; the wretched young man is subjected to the ball of fire which constitutes a Bourbaki discussion. Not only must he understand, but he must also participate. If he is silent, he is simply not invited again.

He must also show a certain quality. The absence of this tendency has stopped many great and valuable mathematicians from joining Bourbaki. During a Congress, the chapters come up in the order of the day, in no particular order, and we never know in advance if we shall be doing only differential topology at this Congress, or if at the next one we shall be doing commutative algebra. No, everything is mixed—I cite the same example, the symbol that could be thought of as the Bourbaki symbol, the ball of wool. Consequently a Bourbaki member is supposed to take an interest in everything he hears. If he is a fanatical algebraist and says "I am interested in algebra and nothing else," fair enough, but he will never be a member of Bourbaki. One has to take an interest in everything at once. Not to be capable of creating in all fields, that is all right. There is no question of asking everyone to be a universal mathematician; this is reserved for a small number of geniuses. But still, one should take an interest in everything, and be able, when the time comes, to write a chapter of the treatise, even if it is not in one's speciality. This is something which has happened to practically every member, and I think most of them have found it extremely beneficial.

In any case, in my personal experience, I believe that if I had not been submitted to this obligation to draft questions I did not know a thing about,

and to manage to pull through, I should never have done a quarter or even a tenth of the mathematics I have done. When one starts to write on questions one does not know and if one is a mathematician, one is forced to put questions to oneself. This is characteristic of the mathematician. Consequently one tries to solve them, and this leads to personal work, independent of Bourbaki, and more or maybe less valuable, but which was born of Bourbaki. So one cannot say that this is a bad system. But there are excellent minds which cannot adapt to this sort of obligation, profound minds which are first-class in their field, but to whom one must not mention other fields. There are unbending algebraists who will never be made to swallow analysis, and analysts for whom the field of quaternions is a monstrosity. These mathematicians may be first-class mathematicians, superior to most Bourbaki members—we admit it freely and I could give you illustrious examples—but they could never be members of Bourbaki.

To return to the guinea pig. When he is invited, we start by looking for this quality of adaptation. Often it is not there, so we wish him luck and he goes on his way. Fortunately one finds from time to time among the youths, this tendency, this appetitie for universal knowledge of mathematics and adaptation to diverse theories. After a very short time, if we find that he gives a good return, he becomes a member without any voting, election, or ceremony. Bourbaki, I repeat, has one rule, which is not to have rules, except for retirement at 50.

To end, I should like to reply to a recent attack on Bourbaki by certain young men of a certain country. Bourbaki is accused of sterilizing mathematical research. I must say that I completely fail to comprehend this, since Bourbaki has no pretension of being a work stimulating to research. I was saying earlier that Bourbaki can only allow himself to write on dead theories, things which have been definitely settled and which only need to be gleaned (except for the unexpected, of course). Actually one must never speak of anything dead in mathematics, because the day after one says it, someone takes this theory, introduces a new idea into it, and it lives again. Rather let us say theories dead at the time of writing, that is to say, nobody has made any significant discoveries in these theories Bourbaki develops for 10, 20, or 50 years, whereas they are in the part judged important and central, serving as tools for research elsewhere. But they are not necessarily stimulants for research. Bourbaki is concerned with giving references and support to anyone who wants to know the essentials in a theory. He is concerned with knowing that when one wants to work, for example, on topological vector spaces there are three or four theorems one has to know: Hahn-Banach, Banach-Steinhaus, the closed graph; it is a question of finding them somewhere. But nobody has the idea of ameliorating the theorems; they are what they are, they are extremely useful (this is the fundamental point) so they are in Bourbaki. This is the important thing. As for stimulating research, if open problems exist in an old theory, obviously they are pointed out, but this is not the aim of Bourbaki.

The aim is, I repeat, to provide worktools, not to give stimulating speeches on the open problems of the new mathematics, because these open problems are in general much farther than Bourbaki can go. This is living mathematics and Bourbaki does not touch living mathematics. He cannot when, by definition, it changes each year. If one wrote a book on that, following Bourbaki's method, i.e., taking eight or ten years to work it out, you can imagine the book after twelve years. It would represent absolutely nothing. It would have to be modified continually and would be like the old Encyclopedia, never finished.

Those are the few explanations I wanted to give you. Now I shall be very happy to answer questions, to add to what I have said.

Answers to questions.

. . . Bourbaki sets off, if you like, from a basic belief, an unprovable metaphysical belief we willingly admit. It is that mathematics is fundamentally simple and that for each mathematical question there is, among all the possible ways of dealing with it, a best way, an optimal way. We can give examples where this is true and examples where we cannot say, because up to now we have not found the optimal method.

I cited, for example, group theory and analytical number theory, which are characteristic. In both one has a quantity of methods, each one more clever than the last. This is splendid and ingenious and of a complexity never before known, but we are sure that this is not the final way to deal with the question. On the other hand, take algebraic number theory. Since Hilbert, it is so systematized that we know there is a right way to handle its questions. We change them sometimes, but in the end, little by little, we manage to find one way which is better than the others. This is only a belief, I repeat, a metaphysical belief.

. . . On foundations we believe in the reality of mathematics, but of course when philosophers attack us with their paradoxes we rush to hide behind formalism and say: "Mathematics is just a combination of meaningless symbols," and then we bring out Chapters 1 and 2 on set theory. Finally we are left in peace to go back to our mathematics and do it as we have always done, with the feeling each mathematician has that he is working with something real. This sensation is probably an illusion, but is very convenient. That is Bourbaki's attitude towards foundations.

* An address before the Roumanian Institute of Mathematics, Bucharest, Oct. 1968.

VISIBLE SHORELINES

F. A. VALENTINE, University of California, Los Angeles

What simple combinatorial geometric condition will enable a collection of swimmers to swim in straight paths to a common point on the shoreline of a convex island? If we assume that the sea is flat, the answer is simple although unexpected. It is true that if every five swimmers can swim linearly to some common point of the shoreline, then all can swim linearly to some common point. As we shall see, if the boundary of the island has four, three or two vertices, then the number "five" can be reduced to "four." If the boundary has at most one vertex, then the corresponding condition on every "three" swimmers suffices. Of course, we are assuming the swimmers are points. To make this precise, the following concepts are used, in which E^n is n-dimensional Euclidean space.

DEFINITION. The interior, closure, boundary, and relative interior of a set C in E^n are denoted by int C, cl C, bd C, intv C respectively. Let $\emptyset$ denote the empty set. A point x exterior to a closed set C in E^n is said to *see a point* $p \in C$ *via the complement* of C if $xp \cap C = \{p\}$, where xp is the closed segment joining x and p.

A *vertex* of a closed convex set C in E^2 is a point of bd C at which there exists more than one line of support to C.

The first result is the following in E^2, the plane.

1. THEOREM 1. *Suppose C is a bounded closed convex set in the plane E^2, and also suppose S is a collection of points in the complement of C.*

(a) *If every five or fewer points in S can see some point of C via the complement of C then there exists at least one point $p \in C$ which all the points of S can see via the the complement of C.*

(b) *If C contains exactly four, three or two vertices, then* (a) *holds with the word "five" replaced by "four."*

(c) *If C contains at most one vertex then the statement* (a) *holds with the word "five" replaced by "three."*

The numbers appearing in these three statements are best in the sense that there exist appropriate sets for which these numbers cannot be reduced.

The idea behind the proof is a combinatorial one of Helly type. To make this clear we need to recall that Helly's theorem [2] on the line states that a collection of closed connected segments on a line E^1 has a nonempty intersection if and only if every pair of segments has a nonempty intersection. (In this statement a segment may degenerate to a point.) The corresponding theorem does

Professor Valentine earned his University of Chicago PhD under G. A. Bliss and L. M. Graves, thesis on the calculus of variations. His main research in convexity has resulted in many articles and his book *Convex Sets*, McGraw-Hill, 1964. He held an instructorship at the Univ. of Tennessee before taking his present position at UCLA. He spent sabbaticals at the Univ. of Chicago, the Univ. of Washington, and the Univ. of Auckland, New Zealand. *Editor.*

not hold for the boundary of a compact convex set C in E^2 with the int $C \neq \emptyset$ as is well known for the circle [1]. However, if we delete one point x from the boundary of C, then a collection of bounded closed arcs of the deleted boundary bd $C \sim \{x\}$ will have a nonempty intersection if and only if every two members of the collection have a nonempty intersection. This follows trivially from the natural 1-1 correspondence that can be set up between E^1 and bd $C \sim \{x\}$. Also it should be noted that if $\mathfrak{N}$ is a collection of bounded closed arcs on bd C and if some s members of $\mathfrak{N}$ have an intersection with bd C which is a nonempty closed arc, then all the members of $\mathfrak{N}$ will have a nonempty intersection if every $s+i$, $i \leqq 2$, members of $\mathfrak{N}$ have a nonempty connected intersection. To see this, observe that s members of $\mathfrak{N}$ reduce the situation to one which is equivalent to Helly's theorem on E^1. This fact will be used to prove Theorem 1 by showing that in (a) s turns out to be 3, in (b) $s=2$, and in (c) $s=1$. Thus, the corresponding "Helly" numbers are "five," "four," and "three" for statements (a), (b), and (c) respectively.

Proof of (a): For $x \in S$, let $C(x)$ denote the set of all points of bd C each of which x can see via the complement of C. Since C is convex, the set $C(x)$ is either an arc or a point of bd C. If for every pair $x_1 \in S$, $x_2 \in S$, the set $C(x_1) \cap C(x_2)$ is connected, then the modification of Helly's theorem [1] as described above and the hypothesis in (a) implies $\bigcap_{x \in S} C(x) \neq \emptyset$, where $\emptyset$ denotes the empty set, and therefore (a) holds.

Hence, to complete the proof, suppose there exist points $x_1 \in S$, $x_2 \in S$ such that $C(x_1) \cap C(x_2)$ is not connected. Since S is convex, this implies that

$$C(x_1) \cap C(x_2) = q \cup \widehat{p_1p_2}, \qquad q \notin \widehat{p_1p_2}, \tag{1}$$

where q is the vertex of C and where $\widehat{p_1p_2}$ is an arc (if $p_1 \neq p_2$) or a point (if $p_1 = p_2$) of the bd C. If $p_1 = p_2$, then the hypothesis in (a) implies that every point of S can see a point p, where $p=q$ or $p=p_1$, so that (a) holds. Therefore, the nontrivial case is that in which $p_1 \neq p_2$. In this case, q, p_1, and p_2 are noncollinear. The three lines determined by q, p_1, p_2 determine seven regions in E^2, as illustrated.

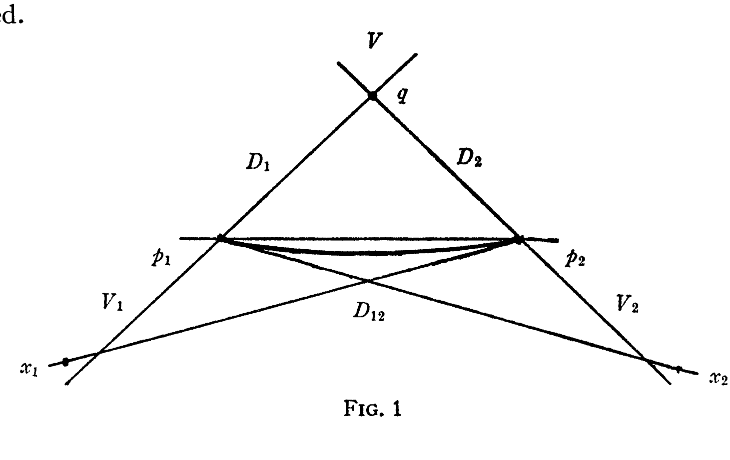


FIG. 1

To be precise, V, V_1, and V_2 are the open unbounded V-shaped regions abutting the triangle at q, p_1, and p_2 respectively. Also D_1, D_2, and D_{12} are the closed three-sided unbounded sets abutting qp_1, qp_2, and p_1p_2 respectively. Since $p_1 \neq p_2$, condition (1) implies that p_1 and p_2 may be reindexed so that $x_1 \in V_1$, $x_2 \in V_2$. (Remember, x_1 and x_2 can each see q and all of arc $\widehat{p_1p_2}$.) If for all $x \in S$, we have $q \in C(x)$, then (a) follows. Hence, suppose there exists a point $x_3 \in S$ such that $q \notin C(x_3)$. This and the hypothesis in (a) implies

$$C(x_3) \cap \widehat{p_1p_2}$$

is a nonempty connected subset of arc $\widehat{p_1p_2}$. It should be observed that this nonempty set is the intersection of three sets $C(x_1)$, $C(x_2)$, and $C(x_3)$, and this corresponds to $s=3$ in our opening introductory paragraph preceding the formal proof. Clearly if $x \in S \sim V$, then x is so situated that $C(x) \cap C(x_3) \cap \widehat{p_1p_2}$ is connected and nonempty. This, together with the hypothesis in (a), and the modification of Helly's theorem [1] mentioned above applied to $\{C(x) \cap C(x_3) \cap \widehat{p_1p_2}, x \in S \sim V\}$ implies that

$$N \equiv \bigcap_{x \in S \sim V} C(x) \neq \emptyset.$$

Now, let $y \in V \cap S$. If both p_1 and p_2 belong to $C(y)$, then the hypothesis in (a), together with the facts $x_1 \in V_1$, $x_2 \in V_2$, implies $C(y) \cap N \subset p_1 \cup p_2$. If there exists a point $y \in V \cap S$ such that $p_1 \notin C(y)$, then $p_2 = C(y) \cap N$; if $p_2 \notin C(y)$, then $p_1 = C(y) \cap N$. These facts together with the hypothesis in (a) imply that if $V \cap S \neq \emptyset$, then $N \cap (\bigcap_{x \in V \cap S} C(x)) \neq \emptyset$, so that

$$\bigcap_{x \in S} C(x) \neq \emptyset. \tag{2}$$

Hence, (a) has been established.

To see that the number five in (a) is best, consider the closed convex set C bounded by a pentagon $A_1B_1C_1D_1E_1$, as illustrated in Fig. 3(α). The set $\{x_i, i=1, \cdots, 5\}$ in Fig. 3(α) is such that every four of its members can see a point of bd C via the complement of C, yet $\bigcap_{i=1}^{5} C(x_i) = \emptyset$.

Proof of (b): (Four vertices). Designate the four vertices of bd C by v_i ($i=1, 2, 3, 4$). If $v_i \in C(x)$ for all $x \in S$ and for a fixed value of i, then (b) follows trivially. Hence, suppose there exist points $x_i \in S$ such that $v_i \notin C(x_i)$ ($i=1, 2, 3, 4$). The hypothesis in (b) implies that

$$M \equiv C(x_1) \cap C(x_2) \cap C(x_3) \cap C(x_4) \neq \emptyset. \tag{3}$$

The set M has no vertices of C. Since each of the $C(x_i)$ ($i=1, 2, 3, 4$) is a closed arc or point of bd C, condition (3) implies that we may reindex the x_i ($i=1, \cdots, 4$) so that

$$M \equiv \text{component of } C(x_1) \cap C(x_2), \qquad C(x_1) \neq C(x_2) \tag{4}$$

or so that

$$M \equiv C(x_1). \tag{5}$$

If (5) holds, then since $C(x_1)$ is connected and contains no vertices of bd C, hypothesis in (b) and the above-mentioned modification of Helly's theorem [1] implies (2) holds, and hence (b) is satisfied. So we consider the case in (4).

If $C(x_1)\cap C(x_2)$ is connected, then since M contains no vertices of bd C Helly's modified theorem can be applied to $\{C(x)\cap C(x_1)\cap C(x_2),\ x\in S\}$ to yield (2), so that (b) holds. Hence, suppose $C(x_1)\cap C(x_2)$ is not connected. Since M contains no vertices of bd C, the nonconnectedness of $C(x_1)\cap C(x_2)$ and the argument that implied (1) implies there exists a vertex, say v_3 (without loss of generality, since $v_1\cup v_2\not\subset C(x_1)\cap C(x_2)$, by construction), such that $v_3\in C(x_1)\cap C(x_2)$, and such that

$$C(x_1)\cap C(x_2) = v_3\cup M,$$

where M is given by (4).

Since M is smooth, consider the two tangents L_1 and L_2 to the extremities of M ($L_1=L_2$ if M is a point). See Figure 2.

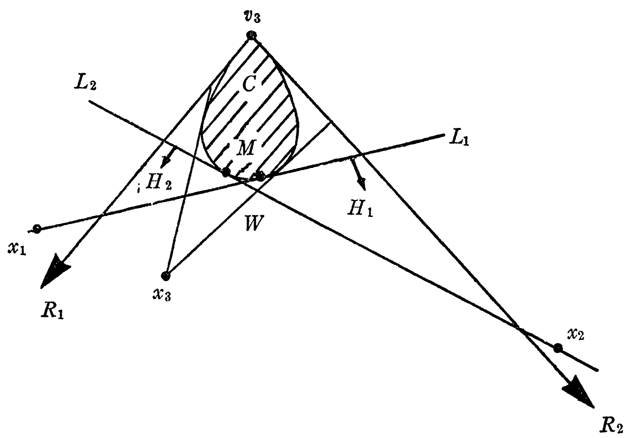


FIG. 2

Also let R_1 and R_2 be the two half-lines of tangency to C at v_3, as illustrated, so that the set C is contained in the closed convex region W bounded by $R_1\cup R_2$. Since $v_3\notin C(x_3)$, the point x_3 is contained in W. Let H_i be the closed half-space bounded by L_i which does not contain C. Furthermore, no points of S exist in the complement of $H_1\cup H_2\cup$convex hull$[M\cup(L_1\cap L_2)]$, otherwise M would contain a vertex of bd C which violates the definition (3). This together with the facts $x_i\in L_i$ ($i=1, 2$), $v_3\notin C(x_3)$ imply that for $x\in S$ the set $C(x)\cap C(x_3)$ is *connected* and nonempty. Hence, since the hypothesis in (b) implies that for all $x\in S$, $y\in S$ we have $C(x)\cap C(y)\cap C(x_3)\neq\varnothing$ (it is also connected). Helly's modified theorem [1] applied to $\{C(x)\cap C(x_3),\ x\in S\}$ implies (2). Thus (b) is true.

If C has exactly three vertices, then the procedure is the same as that above in which, however, we use three sets $C(x_i)$ ($i=1, 2, 3$) with vertices $v_i\notin C(x_i)$.

If C has exactly two vertices v_i ($i=1, 2$), then if (2) fails, there exist sets

$C(x_i)$ $(i=1, 2)$ such that $C(x_1)\cap C(x_2)$ is connected and such that $v_i \notin C(x_1)\cap C(x_2)$, $(i=1, 2)$. This implies $C(x)\cap C(x_1)\cap C(x_2)$ is connected, and Helly's modified theorem [1] applied to $\{C(x)\cap C(x_1)\cap C(x_2),\ x\in S\}$ implies (b).

Figures 3β, 3γ, 3δ illustrate that the number "four" is best in statement (b) since every three of the points (x_1, x_2, x_3, x_4) can see a point of the set C via the complement of C, yet all four cannot see a point of C via the complement of C. The points x_1, x_2, x_3 can see the point 123, etc.

Proof of (c). If bd C has one vertex v_1, then if $v_1 \notin \bigcap_{x\in S} C(x)$ there exists a set $C(x_1)$, $x_1\in S$ such that $v_1 \notin C(x_1)$. Therefore $C(x_1)$ is smooth so that $C(y)\cap C(x_1)$ is connected for all $y\in S$, and Helly's modified theorem applied to $\{C(y)\cap C(x_1),\ y\in S\}$ implies (c).

If bd C is smooth (no vertices) then $C(x)\cap C(y)$ is connected for all $x\in S$, $y\in S$, and (c) follows from Helly's modified theorem. Thus (c) has been proved.

If C is a segment (int $C=\emptyset$), then S can clearly see an endpoint of this segment via the complement of C. The set S consisting of the vertices of a triangle circumscribed about a smooth convex body C shows that the number "three" is best in statement (c).

This completes the proof of Theorem 1.

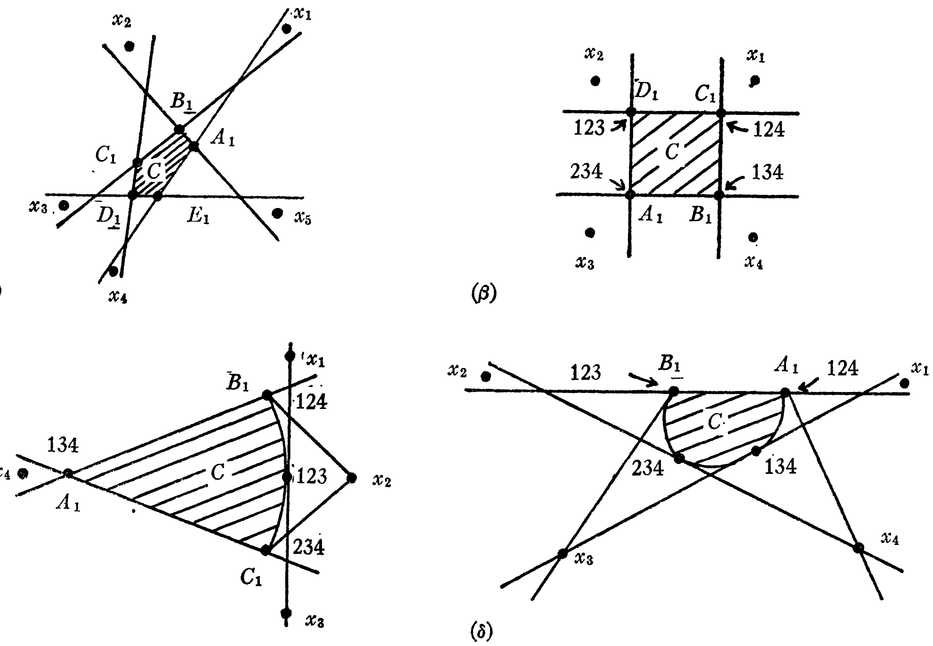


FIG. 3

2. E^n. The story for E^n, $n\geqq 3$ is much more difficult, and we complete the theory only for the case where C is a smooth compact strictly convex set.

THEOREM 2. *Let C be a compact convex set in E^n and suppose C is smooth and*

strictly convex (i.e., C has a unique plane of support at each boundary point and the bd C *contains no line segments). Also let S be a set exterior to C.*

If every $n+1$ or fewer points of S can see some point of bd C *via the complement of C, then there exists at least one point of* bd C *which all of S can see via the complement of C.*

Proof. Since C is smooth, we have int $C \neq \emptyset$. Choose an interior point O of C as origin for a vector space for E^n, and consider the contraction of C, namely $C(\lambda) \equiv \lambda C$, where $0<\lambda<1$. Since C is smooth, the set λC is smooth. The smoothness of C and the hypothesis that every $n+1$ or fewer points of S can see a point of bd C via the complement of C imply that every $n+1$ points of S can be separated from C by some hyperplane H. However, this separation need not be strict so that points of S can lie on H. However, for $\lambda<1$, each set of $n+1$ points of S can be strictly separated from λC, $0<\lambda<1$. Let T be a collection of $n+2$ points selected from $S \cup C(\lambda)$. The separation property mentioned above implies that $T \cap S$ can be strictly separated from $T \cap C(\lambda)$ by a hyperplane $H_1(\lambda)$. Since this holds for every set of $n+2$ points of $S \cup C(\lambda)$, a theorem of Kirchberger [3] (see Appendix) applied to finite subsets of $S \cup C(\lambda)$ implies that each finite subcollection of S can be strictly separated from each finite subcollection of $C(\lambda)$. Hence, by another theorem of Kirchberger [3] (see Appendix), S can be separated from $C(\lambda)$ by a hyperplane $H_1(\lambda)$, not necessarily strictly, however. Let $H(\lambda)$ be the translate of $H_1(\lambda)$ which supports $C(\lambda)$ and such that $H(\lambda)$ and $H_1(\lambda)$ do not contain $C(\lambda)$ between them. Let $x_\lambda \in C(\lambda) \cap H(\lambda)$. Since $\{x_\lambda, 0<\lambda<1\}$ is a bounded set, there exists a sequence $(\lambda_i, i=1, 2, \cdots)$ such that $x_{\lambda_i} \to x \in$ bd C as $i \to \infty$. Also, since $C(\lambda) \to C$ as $\lambda \to 1$, the tangent plane H to C at x must separate S and C. Since $H \cap C = x$, the set S can see x via the complement of C. This completes the proof.

REMARK. The "Helly number" for Theorem 2 in which C is not strictly smooth is still unknown. Since it is possible to see all but one of the vertices of a regular icosahedron, the Helly number is at least 12 for compact convex sets C in E^3. A similar analysis for the dodecahedron will probably yield a higher bound for the best number. The situation in E^n, $n \geqq 3$ is quite involved since the intersection of two different sets of visibility on the boundary of C may have many components. The solution for polyhedra in E^n itself would be of considerable interest.

Appendix

THEOREM (Helly [1], [4]). *Let $\mathfrak{F}$ be a family of compact convex sets in E^n containing at least $n+1$ members. If every $n+1$ members of $\mathfrak{F}$ have a point in common, then all the members of $\mathfrak{F}$ have a point in common.*

THEOREM (Kirchberger, [3], [4]). *Let P and Q be two finite collections of points in E^n. Then P and Q can be strictly separated by a hyperplane if and only if for each subset T of $n+2$ or fewer points of $P \cup Q$ there exists a hyperplane $H(T)$ which strictly separates $T \cap P$ from $T \cap Q$.*

THEOREM (Kirchberger [3], [4]). *Let P and Q be two sets in E^n. If for each set T of $2n+2$ or fewer points of $P\cup Q$ there exists a hyperplane $H(T)$ which separates $T\cap P$ from $T\cap Q$, then there exists a hyperplane which separates P from Q.*

The preparation of this paper was supported in part by NSF Grant Number 8922.

References

1. Hadwiger, Debrunner and Klee, Combinatorial Geometry in the Plane, Holt, Rinehart and Winston, New York, 1964.

2. E. Helly, Über Mengen konvexer Körper mit gemeinshaftlichen Punkten, Jber. Deut. Math. Verein, 32 (1923) 175–176.

3. P. Kirchberger, Über Tschebyschefsche Annäherungsmethoden, Math. Ann., 57 (1903) 509–540.

4. F. A. Valentine, Convex Sets, McGraw-Hill, New York, 1964.

AN APPLICATION OF TAUBERIAN METHODS TO A PROBLEM IN SERIES

DAVID DRASIN, Purdue University

1. Introduction. Let $\{a_n\}$ $(n=1, 2, \cdots)$ be a sequence of positive numbers and

$$\psi(x) = \sum_{n\leq x} a_n.$$

Clearly properties of the function ψ are determined by those of the $\{a_n\}$, and conversely. For example, in a recent note Niven and Zuckerman [8] proved the following result:

THEOREM A. *Let a_n and ψ be defined as above. Furthermore assume*

$$a_1 \geqq a_2 \geqq \cdots \geqq 0. \tag{1.1}$$

Then

$$\lim_{x\to\infty} \{\psi(\sigma x) - \psi(x)\} \tag{1.2}$$

exists for all $\sigma>0$ if and only if

$$\lim na_n \tag{1.3}$$

exists.

I should like to present another proof of Theorem A in which I view (1.1) as a Tauberian condition which allows (1.3) to follow from (1.2). (The derivation of (1.2) from (1.3) is more straightforward.) This proof is self-contained save for a very sketchy justification of Lemma 1. While no new methods are used, I hope that this note will make an important set of ideas—Karamata's

theory of regular varying functions—more available to a larger audience.

Rather than working with series, I consider a nonnegative function $f(t)$, for $1 \leqq t < \infty$. To recover Theorem A from Theorem B below, just choose $f(t) = a_n (n \leqq t < n+1)$ for $n = 1, 2, \cdots$.

THEOREM B. *Let $f(t)$ be nonnegative and monotone decreasing for $1 \leqq t < \infty$, and let $\Psi(x) = \int_1^x f(t)dt$. Then*

$$\lim_{x \to \infty} \{\Psi(\sigma x) - \Psi(x)\} \tag{1.4}$$

exists for each $\sigma > 0$ if and only if

$$\lim_{x \to \infty} xf(x) \tag{1.5}$$

exists.

2. Preliminaries to the derivation of (1.5) from (1.4). I shall reformulate the hypothesis (1.4), and show its Tauberian nature. Thus, let $f(t)$ be positive and nonincreasing on $1 \leqq t < \infty$, and set

$$\Psi(x) = \int_1^x f(t)dt, \qquad \Phi(x) = e^{\Psi(x)}. \tag{2.1}$$

The basic assumption (1.4) implies that corresponding to each $\sigma > 0$ is an $h(\sigma)$ with the property that

$$\frac{\Phi(\sigma x)}{\Phi(x)} \to h(\sigma) \qquad \text{as } x \to \infty. \tag{2.2}$$

Since Φ increases,

(A) *h is monotone increasing.*

Furthermore, it follows from the equation

$$\frac{\Phi(\sigma\tau x)}{\Phi(x)} = \frac{\Phi(\sigma\tau x)}{\Phi(\tau x)} \cdot \frac{\Phi(\tau x)}{\Phi(x)}$$

that h must satisfy

$$h(\sigma\tau) = h(\sigma)h(\tau), \qquad 0 < \sigma, \tau. \tag{B}$$

It is not hard to see, using (A) and (B), that there must be a nonnegative number ρ with

$$h(\sigma) = \sigma^\rho. \tag{2.3}$$

Indeed, the substitutions $\sigma = e^x$ and $\log h = \phi$ transform (B) into

$$\phi(x + y) = \phi(x) + \phi(y), \qquad -\infty < x, y < \infty. \tag{B$'$}$$

Since ϕ increases, it has a point of continuity. It is a very important fact [4, p. 62] that a solution of (B′) which is continuous at a point (and thus continuous

everywhere) must be of the form $\phi(x)=\rho x$ for some real ρ; this is clearly the same as (2.3).

REMARK: The use of this argument in such situations also appears in Feller [2], [3, p. 269].

Equivalently, (2.3) may be written as

$$\Phi(x) = x^{\rho}L(x), \tag{2.4}$$

where, for each $\sigma>0$,

$$L(\sigma x)/L(x) \to 1 \qquad (x \to \infty). \tag{C}$$

Functions satisfying (C) are called *slowly-varying*, and have been the subject of considerable interest (cf. [1], [3], [5], [6], [7], [9]). For present purposes, we need only the following important fact:

LEMMA 1. *A measurable function $L(x)$ is slowly varying if and only if it admits the representation*

$$L(x) = c(x) \exp\left(\int_1^x \{\epsilon(t)/t\}\,dt\right), \tag{2.5}$$

where $\epsilon(t)$ and $c(t)$ are measurable and $\epsilon(t)\to 0$ and $c(t)\to c>0$ as $t\to\infty$.

Proof. The result is standard, so only a sketch will be given here. Since $x^{\rho}L(x)$ increases, it is a simple exercise to verify that the limit in (C) is taken uniformly as $x\to\infty$ in $K^{-1}<\sigma<K$. {In fact, the limit is taken uniformly as $x\to\infty$ in $K^{-1}<\sigma<K$ even if $x^{\rho}L(x)$ is not assumed increasing; for a proof see [7].} Now

$$\int_1^x L(t)dt = \sum_{j=0}^{\infty} \int_{2^{-j-1}x}^{2^{-j}x} L(t)dt, \tag{2.6}$$

and it is not hard to use this uniformity in each of the integrals on the right side of (2.6) to obtain

$$\int_1^x L(t)dt \sim xL(x); \tag{2.7}$$

details appear in [5, p. 25]. An argument due to Karamata [6, pp. 44–45] yields (2.5) from (2.7).

REMARK. The notion of slowly-varying function was due to Karamata, who used (2.7) as his defining property.

Let us now rewrite (2.4), using (2.1) and (2.5):

$$\begin{aligned} \Phi(x) &= \exp\left[\int_1^x f(t)dt\right] = \exp\left[\int_1^x \{tf(t)/t\}\,dt\right] \\ &= c\{1 + o(1)\} \exp\left[\int_1^x \{[\rho + \epsilon(t)]/t\}\,dt\right]. \end{aligned} \tag{2.8}$$

If g is defined by

$$g(t) = tf(t) - [\rho + \epsilon(t)], \tag{2.9}$$

it is clear that (2.8) can be written equivalently as

$$\int_1^x \{g(t)/t\}\, dt \to \log c \qquad (x \to \infty). \tag{2.10}$$

In particular, given $\delta > 0$, there exists an $x_0 = x_0(\delta)$ with the property that if $y > x > x_0$, then

$$\left| \int_x^y \{g(t)/t\}\, dt \right| < \delta. \tag{D}$$

3. Proof that (1.4) implies (1.5). Standard Tauberian arguments applied to (D) will now show that $g(t) \to 0$; i.e.,

$$tf(t) \to \rho \qquad (t \to \infty). \tag{3.1}$$

It is only in this stage that it is convenient to have f decreasing (but see the Remark at the end of this note).

We first establish

$$\limsup_{t \to \infty} tf(t) \leqq \rho. \tag{3.2}$$

If (3.2) were false, there would exist $\eta > 0$ and a sequence $t_n \to \infty$ such that

$$t_n f(t_n) > \rho + \eta \qquad (n = 1, 2, \cdots).$$

By the monotonicity of f,

$$tf(t) \geqq tf(t_n) > (t/t_n)(\rho + \eta), \qquad (1 < t \leqq t_n);$$

hence, if k is given $(0 < k < 1)$, then

$$tf(t) > k(\rho + \eta) \qquad (kt_n \leqq t \leqq t_n).$$

Thus

$$\begin{aligned} \int_{kt_n}^{t_n} \{g(t)/t\}\, dt &\geqq \int_{kt_n}^{t_n} \{[k(\rho + \eta) - \rho - \epsilon(t)]/t\}\, dt \\ &= [k(\rho + \eta) - \rho] \log(1/k) - o(1) \qquad (n \to \infty), \end{aligned}$$

but this contradicts (D), if k is chosen close enough to one.

A similar argument proves $\liminf_{t\to\infty} tf(t) \geqq \rho$, and this together with (3.2) proves (3.1) and hence also (1.5).

4. Derivation of (1.4) from (1.5). The computation in [7] is probably the neatest way of proving the "Abelian" part of Theorem B, but the present proof allows another verification. For suppose

$$\rho = \lim_{t \to \infty} tf(t),$$

and define $\epsilon(t)$ so that the function $g(t)$ introduced in (2.9) vanishes identically. Then

$$\Phi(x) = \exp\left[\int_1^x f(t)dt\right] = x^\rho \exp\left[\int_1^x \{\epsilon(t)/t\}dt\right].$$

If one sets

$$L(x) = \exp\left[\int_1^x \{\epsilon(t)/t\}dt\right],$$

then L clearly satisfies (C), and so Φ satisfies (2.2) with $h(\sigma)=\sigma^\rho$. Note that we do not require that f be decreasing for this part of the proof.

REMARK. It is left as an exercise for the reader to verify that the hypothesis of f decreasing in Theorem B may be weakened to

$$\lim_{\lambda\to 1+} \limsup_{x\to\infty} [f(\lambda x)/f(x)] \leqq 1. \tag{4.1}$$

If f satisfies (4.1), the substitution $F(x)=\log f(x)$ produces a function F which is "slowly decreasing." Slowly decreasing functions have played important roles in Tauberian arguments. The function

$$f(t) = [\mu + \sin t]/t, \qquad \mu > 1$$

fails to satisfy (4.1), and yet

$$\lim_{x\to\infty} \int_x^{\sigma x} f(t)dt = \mu \log \sigma$$

exists for all $\sigma > 0$.

I wish to thank Professors R. Bojanic and D. Shea for several helpful suggestions, and the National Science Foundation for their support (GP-9454).

References

1. D. Drasin, Tauberian theorems and slowly varying functions, Trans. A.M.S., 133 (1968) 333–356.
2. W. Feller, On the classical Tauberian theorems, Arch. Math., 14 (1963) 317–322.
3. ———, An Introduction to Probability Theory and its Applications, Wiley, New York, 1966.
4. C. Goffman, Introduction to Real Analysis, Harper and Row, New York, 1966.
5. G. H. Hardy and W. W. Rogosinski, Notes on Fourier series III, Quart. J. Math. Oxford Ser., 16(1945) 49–58.
6. J. Karamata, Sur un mode de croissance régulière des fonctions, Mathematica (Cluj), 4 (1930) 38–53.
7. J. Korevaar, T. van Aardenne-Ehrenfest, and N. G. de Bruijn, A note on slowly oscillating functions, Nieuw Arch. Wisk., (2) 23 (1949) 77–86.
8. I. Niven and H. S. Zuckerman, On certain sequences, this MONTHLY, 76 (1969) 386–389.
9. D. F. Shea, On a complement to Valiron's Tauberian theorem for the Stieltjes transform, Proc. A.M.S., 21 (1969) 1–9.

LIFE WITHOUT T_2

ALBERT WILANSKY, Lehigh University

1. Introduction. In 1955, Kelley's famous book [4] caused some consternation by the omission of T_2 from several of its definitions. Bourbaki, on the other hand, goes so far as to make T_2 part of the definition of "compact." Dealing with non-Hausdorff spaces has the advantages of greater generality and of working with fewer assumptions. Its (possible) disadvantages are that certain theorems become false when T_2 is dropped, and the extra generality is not paid for by a significant increase in the number of examples covered. We propose in this article to dispose of both of these objections: we shall mention some significant non-Hausdorff spaces, and shall give a number of systematic procedures for extending standard theorems, or modified forms of theorems which, in their usually presented form, deal with Hausdorff spaces.

2. Use retracts instead of closed subspaces. A *retraction* r: $X \to S$ is a continuous map from X to a subspace S satisfying $r(s)=s$ for all $s \in S$. If there exists such a retraction onto S, then S is called a *retract* of X.

Example 2.1. If (X_α) is a family of spaces such that ΠX_α is normal, then each X_α is normal. If "normal" is replaced by "T_4", which means "normal and Hausdorff" this result is proved by the remark that each X_α is homeomorphic with a closed subspace of ΠX_α (the injection map), and T_4 is an F-hereditary property (i.e., inherited by closed subspaces). Without the T_2 assumption, the injection of X_α into the product need not be closed, so this argument fails. However, the projection P_α from the product onto X_α is a retraction (if we identify X_α with its injection into the product), and normal is an r-hereditary property. (A retract of a normal space is normal.) Thus if ΠX_α is normal, so is each X_α.

Example 2.2. The list of r-hereditary properties is very long, including all F-hereditary T_2 properties. It also includes all C-hereditary properties (properties preserved by taking a continuous image). Thus if ΠX_α is connected, so is each X_α. But it also includes properties which are neither F- nor C-hereditary; one such is that of being a Baire space. Let X be the space obtained by deleting the real irrationals from the complex plane; X is a Baire space but its closed subspace, the real rationals, is not. But Baire is r-hereditary, so if ΠX_α is a Baire space, so is each X_α.

Example 2.3. The classical theorem: "*If X is a Hausdorff space, the diagonal in $X \times X$ is closed*" fails for more general spaces. It may be replaced by the following generalization: The graph of continuous f: $X \to Y$ is a retract of $X \times Y$. (Proof: Consider $(x, y) \to (x, fx)$.) Thus we obtain: If X and Y are T_2 spaces and f: $X \to Y$ is continuous, then its graph is closed. The first mentioned result is the application of this to i: $X \to X$.

Using retractions is also a convenient way of proving T_2 theorems.

Example 2.4. (See [2], Lemma 6.11.) If D is a dense subspace of a T_2 space X, if ϕ: $X \to Y$ is continuous, and if $\phi|D$ is a homeomorphism, then $\phi[\tilde{D}] \cap \phi[D] = \varnothing$, i.e., ϕ carries $X \backslash D$ and D into disjoint sets. Proof: Let $W = \phi^{-1}[\phi D]$

and $r=(\phi^{-1}|\phi[D])\circ(\phi|W)$. Then r is a retraction of W onto D, hence D is closed and dense, so $D=W$.

Example 2.5. Let each X_α be a compact T_2 space and $X=\Pi X_\alpha$. Extend each projection P_α: $X\to X_\alpha$ to r_α: $\beta X\to X_\alpha$. Define r: $\beta X\to X$ as the map whose αth coordinate is r_α. Then r is a retraction, hence X is closed and dense, so $X=\beta X$ and X is compact. This is (a case of) Tychonoff's product theorem.

3. Use closed graph instead of continuous. We call f: $X\to Y$ a CG function if it has closed graph, i.e., the graph of f is a closed subset of $X\times Y$. Since a continuous map between T_2 spaces is CG, any theorem in which CG is a hypothesis generalizes the corresponding theorem for continuous maps between T_2 spaces. Moreover, the substitution restores the truth of some such theorems.

Example 3.1. If X is compact, Y is T_2, and f: $X\to Y$ is continuous, then f is closed (i.e., f preserves closed sets). This fails if Y is not T_2, but we have the generalization: If X is compact and f: $X\to Y$ is CG, then f is closed. (Proved in [**8**], Section 11.1, Fact ii.) The first mentioned result is a special case.

The casuality in dropping T_2 is the uniqueness of limits of convergent sequences, nets, filters. However, this uniqueness is recaptured by the CG assumption, as shown in the next example.

Example 3.2. Let X and Y be real linear topological spaces, and f: $X\to Y$ a CG additive map. Then f is linear. Proof: As usual, $f(rx)=rf(x)$ for all x, and for rational r. For any real t, let $r\to t$ through rational values. Then $rx\to tx$ and $f(rx)=rf(x)\to tf(x)$. Since $[rx, f(rx)]\in G$ the graph of f, and G is closed, we have $[tx, tf(x)]\in G$, hence $tf(x)=f(tx)$, completing the proof. The corresponding theorem for continuous maps and T_2 spaces (which is, of course, a special case) may be proved by the same argument except that uniqueness of limits is used in the last step instead of CG.

There are ways in which CG functions are better behaved than continuous ones. The following two remarks fail for continuous functions: If f is a CG bijection, f^{-1} is also CG. (Proof: It has the same graph as f.) If f: $X\to Y$ is CG, then f remains CG when the topologies of X and Y are enlarged. (Proof: A closed set remains closed when the topology is enlarged.) The latter remark is the Closed Graph Lemma which has numerous applications in functional analysis. Moreover, one of the most famous operators in all mathematics is CG and not continuous. This is D: $C^{(1)}\to C$, the differentiation operator. The relevant facts are that if (a) $f_n\to f$ uniformly and (b) $f_n'\to g$ uniformly, then $g=f'$, i.e., (c) $f_n'\to f'$ uniformly; and that (a) alone does not imply (c).

Using CG functions is sometimes a good way to prove T_2 theorems.

Example 3.3. Let X and Y be homeomorphic locally compact T_2 spaces, and let X_1 and Y_1 be the compact T_2 spaces obtained by adjoining one point to X, Y respectively. Then X_1 and Y_1 are homeomorphic, with the added points corresponding to each other. This standard result on uniqueness of one-point compactification may be proved thus: Extend the given homeomorphism to a bijection f: $X_1\to Y_1$ in the obvious way. It is easy to check that f is CG. Hence, by

the "Little closed graph theorem" ([8], Section 11.1, Fact iv), f is continuous. Similarly f^{-1} is continuous.

We remark that a category in which the morphisms are CG functions must be restricted; for example the identity map is not CG unless its domain is Hausdorff and the composition of CG functions need not be CG. (See [8], Section 11.2, Problem 17.)

4. Use locally compact instead of compact. Our intention here is that a locally compact space is one which has a local base of compact neighborhoods of each point. A compact space need not be locally compact; (see [9]) but this need disturb us no more than the fact that a connected space need not be locally connected.

Example 4.1. Consulting [4], p. 223, Theorem 5, we find the standard theorem which states that if F is a family of continuous maps from a regular or a Hausdorff space X to a topological space Y, and C is the compact open topology, then C is the smallest topology which is jointly continuous on compact sets. There also we find that half of the theorem remains true without separation, namely, for arbitrary X, the topology C is smaller than every topology which is jointly continuous on compact sets. We add the remark that the other half of the theorem can be similarly reclaimed in the form: For arbitrary X, the topology C is jointly continuous on locally compact sets. (The proof given in [4] can be easily modified to cover this.) The standard theorem follows immediately for any space X in which compact subsets are locally compact.

5. Assume regular instead of T_2. Students of [4] will be familiar with this suggestion. For regular spaces compact implies locally compact, locally compact implies second category, compact implies normal and completely regular. All uniform spaces (in particular topological groups and vector spaces) are regular.

Example 5.1. Let X be regular, then X^+ is regular if and only if X is locally compact. (X^+ is the one point compactification of X.) The standard theorem ([4], p. 150), that if X is T_2, then X^+ is T_2 if and only if X is locally compact, follows immediately. (For example, if X is locally compact T_2, then X^+ is regular and T_1, hence T_3, hence T_2.)

Example 5.2. A regular second countable space is semimetrizable. (A semimetric is like a metric except that $d(x, y)=0$ for $x \neq y$ is allowed.) Also (Nagata-Smirnov Theorem) a regular space is semimetrizable if and only if it has a σ-locally finite base. The usual proofs may be easily modified by replacing embedding by the weak topology of a countable family of maps into a metric space. (These results fail for normal spaces, however, since a normal space need not be regular.)

6. Assume KC instead of T_2. A KC space is one in which all compact sets are closed. (See [9].)

Example 6.1. Let T be a compact T_2 topology for a set X. Then T is maximal among compact topologies for X. This result (with precisely the same proof) holds if T_2 is replaced by KC. But now we have the advantage that the converse

holds, i.e., a compact topology is maximal among compact topologies if and only if it is KC. (See [6], Theorem 2.)

Example 6.2. In the absence of T_2 one may have compact sets whose closures and intersections are not compact. See [3] for some extreme pathology. Those who are made uncomfortable by such examples may work in KC spaces since such spaces do not allow them.

7. Assume symmetric instead of T_2. A symmetric space is one in which $x\in\overline{\{y\}}$ implies $y\in\overline{\{x\}}$. For symmetric spaces T_0 implies T_1, normal implies regular, normal T_0 implies T_4.

8. Modify conclusions of theorems. This is done in Examples 5.1, 5.2.

Example 8.1. The result that a finite dimensional Hausdorff linear space has a unique topology is replaced by: Every finite dimensional linear topological space has its topology uniquely determined by the closure of $\{0\}$. (See [8], Section 10.6, Theorem 1.)

Example 8.2. The Mackey-Arens Theorem goes through unchanged without separation assumptions. (See [5], Theorem 18.8.) The crucial step that a certain compact set is closed ([5], p. 143, line 7) is carried out in an associated Hausdorff space.

9. Modify definitions. Neighborhoods of ∞ in the one point compactification X^+ must be defined as complements of compact closed sets (rather than just compact sets). Also, for certain purposes σ-compact and hemi-compact must be replaced by σ-(compact closed) and hemi-(compact closed); for example, $\{\infty\}$ is a G_δ in X^+ if and only if X is σ-(compact closed), and X^+ is first countable at ∞ if and only if X is hemi-(compact closed). For KC spaces no such modification is needed since compact sets are already closed. This remark applies to the rest of this section also. Paracompactness may successfully ask for regularity rather than T_2 in its definition. For example the famous theorem of A. H. Stone says that a semimetric space is paracompact. (See [4] p. 160, Corollary 35.)

Example 9.1. Let P be a topological property. A *P-generated space X* is one in which a subset S is closed if $S\cap K$ is closed in K for all sets K which have property P. For P = compact, a P-generated T_2 space is called a *k-space*. Steenrod [7] makes a compelling case for doing topology in the category of Hausdorff k-spaces and continuous maps. He points out that a regular open subspace of a k-space is a k-space; actually it turns out that every open subspace of a k-space is a k-space. This is immediate from [1], Theorems 11.9.4 and 6.2.1.

The category of $KC.k$-spaces and continuous maps is convenient for some purposes; for example (see [9], Theorem 5), it admits one point compactification (not as a functor, of course); but it is not offered as a serious alternative since, for example the product of a compact KC space with itself is not KC unless the space is T_2. In [4], the definition of k-space for non-T_2 spaces is obtained by taking P = compact closed. The following advantages of retaining the original definition, P = compact, were pointed out to me by J. W. Taylor: a first

countable or locally compact space must be a k-space in this sense (compare [4], p. 231, Theorem 13); a space in which each compact subset is locally compact is a k-space if and only if it is a quotient of a locally compact space (compare [1], Theorem 11.9.4); a quotient of any k-space is a k-space.

10. Specialize when necessary. In some situations T_2 may be needed or convenient. In such cases one specializes; but this is no reason to use T_2 as a blanket assumption from the beginning, any more than prospective use of the Čech compactification justifies dealing only with completely regular spaces from the beginning.

11. Examples. Some examples of useful non-Hausdorff topologies are: the topology induced by a non-separating family of seminorms on a vector space, the one point compactification (see [9]), the Zariski topology used by algebraic geometers.

References

1. J. Dugundji, Topology, Allyn and Bacon, Boston, 1966.
2. L. Gillman and M. Jerison, Rings of Continuous Functions, Van Nostrand, Princeton, N. J., 1960.
3. H. Herrlich, On intersections of compact sets, Duke Math. J., 35 (1968) 439–440.
4. J. L. Kelley, General Topology, Van Nostrand Princeton, N. J., 1955.
5. J. L. Kelley, I. Namioka *et al.*, Linear Topological spaces, Van Nostrand, Princeton, N. J., 1963.
6. N. Levine, When are compact and closed equivalent? this MONTHLY, 72 (1965) 41–44.
7. N. Steenrod, A convenient category of topological spaces, Michigan Math. J., 14 (1967) 133–152.
8. A. Wilansky, Functional Analysis, Blaisdell, Waltham, Mass., 1964.
9. ———, Between T_1 and T_2, this MONTHLY, 74 (1967) 261–266.

MATHEMATICAL NOTES

EDITED BY DAVID DRASIN

Manuscripts for this Department should be sent to David Drasin, Division of Mathematical Sciences, Purdue University, Lafayette, IN 47907.

ON DIVIDING A SQUARE INTO TRIANGLES

PAUL MONSKY, Brandeis University and Kyoto University

Sometime ago in this MONTHLY, Fred Richman and John Thomas [1] asked the following puzzling question:

Can a square S be divided into an odd number of nonoverlapping triangles T_i, all of the same area?

In [2], the answer was shown to be no, provided $S=[0, 1]\times[0, 1]$ and the coordinates of the vertices of the T_i are rational numbers with odd denominators. In this note we shall show that the answer is always no. In fact we shall prove the following more general result.

Suppose that $S=[0,1]\times[0,1]$ *is divided into* m *nonoverlapping triangles* T_i; *let* $a_i=\text{area } T_i$. *Then there is a polynomial* f *with integer co-efficients such that* $f(a_1,\cdots,a_m)=1/2$.

There are two parts to the proof: one combinatorial, the other valuation theoretic. The combinatorial argument generalizes an argument made in [2]. By itself it may be made to prove the desired result when the vertices of the T_i all have rational coordinates. But to handle the case of arbitrary vertices, it becomes necessary to argue with "congruences mod 2 in the reals." This is where valuation theory comes in; we make use of absolute values on the reals extending the 2-adic absolute value of the rationals. In the course of the proof the theorem of the extension of valuations plays a remarkable and unexpected role.

We begin with the combinatorial argument. Let R be a region in the plane bounded by a simple closed polygon. Suppose R is divided into m nonoverlapping triangles T_i. By a *vertex* we shall mean a vertex of some T_i, by a *face* a face of some T_i or of R. Two vertices are called *adjacent* if they are in the same face and the line segment joining them contains no other vertices. A *basic segment* is a line segment joining two adjacent vertices. Note that the boundary of each T_i is a union of nonoverlapping basic segments; the same is true of the boundary of R. Suppose now that the vertices are divided into three disjoint sets, $\mathfrak{A}$, $\mathfrak{B}$, and $\mathfrak{C}$. We shall say that a face or a basic segment is of *type* $\mathfrak{AB}$ if it has one endpoint in $\mathfrak{A}$ and one in $\mathfrak{B}$.

LEMMA. *Suppose that no face contains vertices of all three types and that* R *has an odd number of faces of type* $\mathfrak{AB}$. *Then some* T_i *has vertices of all three types.*

To prove the lemma note the following. A face of type $\mathfrak{AB}$ contains an odd number of basic segments of type $\mathfrak{AB}$, while a face not of type $\mathfrak{AB}$ contains an even number of basic segments of type $\mathfrak{AB}$. (Use the fact that a face contains vertices of at most two types to prove this.) Suppose that no T_i has vertices of all three types. Then each T_i has either 0 or 2 faces of type $\mathfrak{AB}$. Hence the boundary of T_i contains an even number of basic segments of type $\mathfrak{AB}$. Similarly, the boundary of R contains an odd number of basic segments of type $\mathfrak{AB}$. But this is impossible; in an obvious sense the boundary of R is congruent to the sum of the boundaries of the T_i modulo 2.

We now come to the valuation theoretic part of the proof and need to introduce some further terminology. Let K be a field. By an *ultranorm* (sometimes called a *non-Archimidean absolute value*) on K we mean a function $\|\ \|$ from K to the nonnegative real numbers satisfying:

$$\|xy\| = \|x\|\cdot\|y\| \tag{1}$$

$$\|x+y\| \leqq \max(\|x\|,\|y\|) \tag{2}$$

$$\|x\| = 0 \Leftrightarrow x = 0. \tag{3}$$

We can easily prove that $\|1\|=\|-1\|=1$, and that equality holds in equation (2) unless $\|x\|=\|y\|$.

As an example let K be the field of rational numbers. Any $x \neq 0$ in K may be written as $2^t(r/s)$, where r and s are odd integers and t is an integer. Set $\|x\| = (1/2)^t$. In this way we get an ultranorm on the rationals in which $\|2\| < 1$. This ultranorm is known as the *2-adic absolute value.* The more general fact that we shall need is this: *There is an ultranorm on the field of real numbers (or more generally on any extension of the rational numbers) such that* $\|2\| < 1$. This follows from the theorem of the extension of valuations whose proof may be found in many places; for example see [**3**].

Granting the above facts we may argue as follows. Choose an ultranorm on the reals for which $\|2\| < 1$. Divide the points of the plane into three sets in the following way:

$$(1) \qquad (x, y) \text{ is in } \mathfrak{A} \text{ if } \|x\| < 1 \text{ and } \|y\| < 1,$$

$$(2) \qquad (x, y) \text{ is in } \mathfrak{B} \text{ if } \|x\| \geqq 1 \text{ and } \|x\| \geqq \|y\|,$$

$$(3) \qquad (x, y) \text{ is in } \mathfrak{C} \text{ if } \|y\| \geqq 1 \text{ and } \|y\| > \|x\|.$$

Suppose now that $P = (x, y)$ and $P' = (x', y')$ are points and that P' is a translate of P by a point of type $\mathfrak{A}$; in other words, that both $\|x' - x\| < 1$ and $\|y' - y\| < 1$. *Then P and P' have the same type.* If P is of type $\mathfrak{A}$ this is obvious. If P is of type $\mathfrak{B}$, then $\|x'\| = \|x\| \geqq 1$, while $\|y'\| \leqq \max(1, \|y\|) \leqq \|x\| = \|x'\|$; so P' is of type $\mathfrak{B}$ too. If P is of type $\mathfrak{C}$ the argument is similar.

It is now easy to see that a line L cannot contain points of all three types. For by translating a point of type $\mathfrak{A}$ on L to the origin we may assume that $(0, 0)$ is on L. Let (x, y) and (x', y') be points of L of types $\mathfrak{B}$ and $\mathfrak{C}$. Then $\|x\| \geqq \|y\|$, $\|y'\| > \|x'\|$, and $\|xy'\| > \|x'y\|$. This is absurd as $xy' = x'y$. Note also that if a triangle $\boldsymbol{T}$ has vertices of all three types then $\|\text{area } \boldsymbol{T}\| > 1$. For we may assume that the vertex of $\boldsymbol{T}$ of type $\mathfrak{A}$ is $(0, 0)$. Let (x, y) and (x', y') be the vertices of types $\mathfrak{B}$ and $\mathfrak{C}$. Then area $\boldsymbol{T}$, up to sign, is equal to $\frac{1}{2}(xy' - x'y)$. But $\|xy'\| > \|x'y\|$. So $\|\text{area } \boldsymbol{T}\| = \|\frac{1}{2}\| \, \|xy'\| = \|\frac{1}{2}\| \cdot \|x\| \cdot \|y'\| > 1$.

Suppose now that $\boldsymbol{S} = [0, 1] \times [0, 1]$ is divided into m nonoverlapping triangles $\boldsymbol{T}_i$ each of area $1/m$. Obviously $\boldsymbol{S}$ has exactly one face of type $\mathfrak{A}\mathfrak{B}$; by the lemma, some $\boldsymbol{T}_i$ has vertices of all three types. By the paragraph above, $\|\text{area } \boldsymbol{T}_i\| = \|1/m\| > 1$. So m is even. (Note that if all vertices have rational coordinates, then we can argue directly with the 2-adic absolute value of the rationals, and avoid the theorem of the extension of valuations; this is essentially what was done in [**2**].)

Finally, we indicate the proof of the more general theorem mentioned at the beginning of this paper. Let $\boldsymbol{A}$ be the ring $\boldsymbol{Z}[a_1, \cdots, a_m]$. If 2 generates the unit ideal in $\boldsymbol{A}$, then $1 = 2f(a_1, \cdots, a_m)$ and we are done. If $2\boldsymbol{A} \neq \boldsymbol{A}$, then 2 is contained in a height 1 prime ideal $\boldsymbol{P}$ of $\boldsymbol{A}$. The integral closure of the local ring of $\boldsymbol{P}$ on $\boldsymbol{A}$ is a discrete valuation ring. This ring gives rise to an ultranorm on the quotient field of $\boldsymbol{A}$ such that $\|a_i\| \leqq 1$, while $\|2\| < 1$. Extend this ultranorm to the reals, and use it to subdivide the plane into points of three types as above. Then, as above, some $\boldsymbol{T}_i$ has vertices of all three types, and $\|a_i\| = \|\text{area } \boldsymbol{T}_i\| > 1$,

a contradiction. (By using valuation rings instead of ultranorms we could simplify the proof a little.)

The above proof is not so wildly nonconstructive as it first appears. For the entire argument is carried out in the field generated by the coordinates of the vertices. So it is only necessary to extend our ultranorm from Q to this finitely generated field, not to the entire field of real numbers.

References

1. Fred Richman and John Thomas, Problem 5479, this MONTHLY, 74 (1967) 329.
2. John Thomas, A dissection problem, Math. Mag., 41 (1968) 187–190.
3. Serge Lang, Algebra, Addison-Wesley, Reading, Mass., 1965, p. 299.

SYLVESTER'S PROBLEM ON COLLINEAR POINTS AND A RELATIVE

G. D. CHAKERIAN, University of California, Davis

In 1893, J. J. Sylvester posed the following question in the Educational Times: *Given a finite set of points in the plane, not all lying on a straight line, must there be a line containing exactly two of the points?*

This question was not settled until after 1930, when T. Gallai proved the affirmative answer. Many proofs have been given since that time, some sparked by a restatement of the problem by P. Erdös [3] in this MONTHLY. V. C. Williams [9] recently gave a proof in this MONTHLY. Extensive references to the literature which has grown up around the problem can be found in [2], [4], and [5].

If we settle upon the projective plane as an appropriate setting and state the problem in its dual form, we obtain: *Given n lines in the projective plane, not all concurrent, must there be a point lying on exactly two of the lines?*

Thinking of the projective plane as a sphere with antipodes identified, and then immediately observing that the identification contributes nothing to the problem, we obtain the following equivalent formulation of Sylvester's problem: *Given n great circles on a sphere, not all concurrent, must there be a point lying on exactly two of the great circles?*

In his exquisite paper on zonohedra, H. S. M. Coxeter [1] observes in passing that in this context the positive answer to Sylvester's question follows immediately from the fact that there exists no map on the sphere (each of whose countries has at least 3 sides) such that all vertices have valence 6 or more. For a set of n great circles such that no point lies on exactly two circles, would form a map on the sphere such that each vertex has valence at least 6.

Here is a short proof of this fact about maps. Suppose we are given a map on the sphere such that each face (country) has 3 or more sides. Suppose each vertex had valence at least 6 (the valence of a vertex is the number of edges emanating from that vertex). If V denotes the total number of vertices, E the number of edges, and F the number of faces, then we have Euler's relation,

$$V - E + F = 2. \tag{1}$$

Let F_k denote the number of k-sided faces of the map. Counting the number of edges around each face and summing, we obtain

$$2E = 3F_3 + 4F_4 + \cdots \geqq 3F, \tag{2}$$

where we have used the assumption that $F_2=0$. Let V_k be the number of vertices of degree k. Counting the number of edges around each vertex and summing, we get

$$2E = 6V_6 + 7V_7 + \cdots \geqq 6V, \tag{3}$$

where we use the assumption that $V_k=0$ for $k\leqq 5$. Now (1), (2), and (3) imply

$$6(E+2) = 6(V+F) \leqq 2E + 4E = 6E,$$

a contradiction. Hence there must be vertices of valence less than 6.

The above solution of Sylvester's problem, which does not appear to be as well known as some others, was offered by N. E. Steenrod in response to the problem of Erdös [3]. This approach is used in [4] to obtain a number of other results on configurations of points and lines.

The relative of Sylvester's problem referred to in the title is the following: *Given a finite set of points in the plane, each colored either red or blue and not all collinear, must there be a "monochromatic line"? That is, must there be a line, containing at least two points of the given set, such that all points of the set on that line are the same color?*

T. S. Motzkin and M. Rabin ([6], [7]) have proved this is so. The problem is attributed to R. Graham and D. Newman. We shall give another proof, similar in spirit to Steenrod's solution of Sylvester's problem, using the following topological lemma of A. L. Cauchy.

CAUCHY'S LEMMA. *Given a map on a sphere such that each face has at least* 3 *sides, it is not possible to label the edges with* $+$*'s and* $-$*'s in such a way that the number of sign changes around each vertex is at least* 4.

To obtain "the number of sign changes around a vertex" we run once around that vertex, counting the number of times we have a sign change (from $+$ to $-$ or *vice versa*) as we traverse the edges emanating from that vertex. In other words, a "sign change" is simply a pair of adjacent edges labeled with opposite signs. Cauchy's lemma is valid even if some edges are allowed to remain unlabeled. The lemma is needed in that more general form in order to prove Cauchy's famous rigidity theorem: *a convex polyhedron cannot be deformed into another noncongruent convex polyhedron without stretching or tearing its surface.*

For a proof of this theorem, and Cauchy's lemma in its more general form, the reader may consult [8]. We now give a proof of Cauchy's lemma as stated above.

Suppose we have a map on a sphere such that each country has at least 3 sides. Suppose the edges have been labeled in such a fashion that the number

of sign changes around each vertex is at least 4. Let N be the sum of all the sign changes around all the vertices of our map. In other words, run around each vertex once, counting the sign changes, and sum over all vertices—the result is N. Since the number of sign changes around each vertex is assumed to be at least 4, we have

$$N \geqq 4V. \tag{4}$$

A bit of reflection shows that if we run around the edges of each face, counting the number of sign changes as we go once around that face, and sum over all faces, then the result is also N. (Note that what we are doing here is quite analogous to what we did in deriving (2) and (3). But instead of counting edges, we are counting pairs of oppositely labeled adjacent edges.) Since the total number of sign changes, as we run around a k-sided face, is even and at most k, we have,

$$N \leqq 2F_3 + 4F_4 + 4F_5 + 6F_6 + 6F_7 + \cdots. \tag{5}$$

Using (1) and (2) (which are still valid for our map) and then (5) and (4), we have,

$$\begin{aligned} 4V - 8 = 4E - 4F &= 2(3F_3 + 4F_4 + \cdots) - 4(F_3 + F_4 + \cdots) \\ &= 2F_3 + 4F_4 + 6F_5 + \cdots \geqq N \geqq 4V, \end{aligned}$$

a contradiction, and the proof is complete.

In order to answer the question of Graham and Newman, we begin, as with Sylvester's problem, by dualizing and transferring the problem onto a sphere. Then what we want to prove is: *Given n great circles on a sphere, each colored either red or blue and not all concurrent, there is a "monochromatic point." That is, there is a point lying on at least two of the circles and such that all circles through that point are the same color.*

To prove this, label each edge of the resulting map on the sphere with a "+" or a "−," depending on whether it lies on a red circle or a blue circle. If there were no monochromatic points, each vertex of the map would have both red circles and blue circles passing through it; hence there would clearly be at least 4 sign changes around each vertex. Since each face of our map has at least 3 edges, Cauchy's lemma shows this is impossible, so there exist monochromatic points.

References

1. H. S. M. Coxeter, The classification of zonohedra by means of projective diagrams, J. Math. Pures Appl., 41 (1962) 137–156.

2. D. W. Crowe and T. A. McKee, Sylvester's problem on collinear points, Math. Mag., 41 (1968) 30–34.

3. P. Erdös, Problem No. 4065, this MONTHLY, 51 (1944) 169.

4. L. M. Kelly and W. O. J. Moser, On the number of ordinary lines determined by n points, Canad. J. Math., 10 (1958) 210–219.

5. T. S. Motzkin, The lines and planes connecting the points of a finite set, Trans. Amer. Math. Soc., 70 (1951) 451–464.

6. ———, Nonmixed connecting lines, Notices, Amer. Math. Soc., 14 (1967) 837.

7. ——— and M. Rabin, Nonmixed connecting lines, J. Comb. Th., to appear.

8. L. A. Lyusternik, Convex Figures and Polyhedra, Russian, English translation by T. Jefferson Smith, Dover, New York, 1963.

9. V. C. Williams, A proof of Sylvester's theorem on collinear points, this MONTHLY, 75 (1968) 980–982.

ON NORMALIZED SCHAUDER BASES

T. A. COOK, University of Massachusetts

In 1948, Karlin [1] proved that in Hilbert space a normalized Schauder basis is an orthonormal basis. His proof is somewhat involved and in this note an extremely simple proof is offered.

A *Schauder basis* in a Banach space E is a sequence $\{x_n\}$ of vectors in E together with a sequence of continuous linear functionals $\{f_n\}$ such that $f_k(x_j)=\delta_{k,j}$ and such that each x in E has the representation

$$x = \sum_{n=1}^{\infty} f_n(x)x_n.$$

It is understood that the partial sums of the above series converge in norm to x. (The reader is reminded that this is equivalent to saying that each $x\in E$ has a unique representation as $x=\sum_{i=1}^{\infty} a_i x_i$, see, e.g., [2, pp. 86–89].) A Schauder basis is called *normalized* provided $\|x_n\|=\|f_n\|=1$ for each n.

THEOREM. *Each normalized Schauder basis in Hilbert space is an orthonormal basis.*

Proof. If $k\neq j$, we must show that x_k is orthogonal to x_j; i.e., $(x_k, x_j)=0$, where (x, y) is the inner product in the Hilbert space. Suppose for a particular k and j that $(x_k, x_j)\neq 0$. Then there is a unit vector e in the span of x_k and x_j which is orthogonal to x_k. Consequently, $|(x_j, e)|>0$. Now, $x_j=(x_j, e)e+(x_j, x_k)x_k$ and $1=\|x_j\|^2=|(x_j, e)|^2+|(x_j, x_k)|^2$. Thus $1>|(x_j, e)|$. Hence, $1=|f_j(x_j)|=|(x_j, e)|\cdot|f_j(e)|$ and $|f_j(e)|=1/|(x_j, e)|>1$. But this denies that $\|f_j\|=1$. Consequently, x_k must be orthogonal to x_j.

This research was supported in part by the National Science Foundation Grant GP-5952.

References

1. S. Karlin, Bases in Banach spaces, Duke Math. J., 15 (1948) 971–985.

2. A. Wilansky, Functional Analysis, Blaisdell, New York, 1964.

AN IDENTITY

J. SHEEHAN, University of Waterloo

Let n be a positive integer. Let $[j]=(1,^{j_1} 2^{j_2}, \cdots, n^{j_n})$ denote a partition of n, where $1j_1+2j_2+\cdots+nj_n=n$.

THEOREM.

$$\frac{1}{n}=\sum_{[j]}(-1)^{j_1+j_2+\cdots+j_n+1}\frac{(j_1+j_2+\cdots+j_n-1)!}{j_1!j_2!\cdots j_n!},$$

where the summation is over all partitions of n.

Proof. Let $l(x^n)[\cdots]$ mean "the coefficient of x^n in $[\cdots]$." Then

$$\begin{aligned}\frac{1}{n} &= l(x^n)\left[x+\frac{x^2}{2}+\cdots+\frac{x^n}{n}\right]\\ &= l(x^n)[\log(1-x)^{-1}]\\ &= l(x^n)[\log 1+(x+x^2+\cdots+x^n)]\\ &= l(x^n)\Big[-(x+x^2+\cdots+x^n)+(-1)^2\left(\frac{x+x^2+\cdots+x^n}{2}\right)^2+\cdots\\ &\qquad\qquad+(-1)^n\left(\frac{x+x^2+\cdots+x^n}{n}\right)^n\Big]\\ &= \sum_{[j]}(-1)^{j_1+j_2+\cdots+j_n+1}\frac{(j_1+j_2+\cdots+j_n-1)!}{j_1!j_2!\cdots j_n!}.\end{aligned}$$

(Note that 0! is defined to be equal to 1.)

I am indebted to the referee for pointing out to me that the above identity is also a consequence of inverse relations for Bell polynomials [1, p. 174].

Reference

1. J. Riordan, An introduction to combinatorial analysis, a Wiley publication in Mathematical Statistics, Chapman and Hall, Ltd., London, 1958.

ON THE CIRCLE GROUP OF A NILPOTENT RING

R. L. KRUSE, Sandia Laboratories, Albuquerque

It is well known [1] that the Jacobson radical of a ring forms a group under the circle composition

$$x \circ y = x + y + xy.$$

It would be of interest to characterize the groups which can occur as circle groups. In the finite case it is sufficient to study p-groups. To see this, observe that the primary decomposition of the additive group of a finite ring induces a

ring direct sum; hence the circle group of a finite radical ring is the direct product of its Sylow subgroups.

A fundamental property of a finite p-group G is that G possesses a central series, that is, a series of normal subgroups

$$G = Z_0 \supset Z_1 \supset \cdots \supset Z_c = 1$$

such that Z_{i-1}/Z_i is in the center of G/Z_i for $1 \leqq i \leqq c$. The length c of the shortest such series is called the *class* of G. Clearly each group G of class 1—that is, each Abelian group—occurs as the circle group of the ring with additive group G and all products 0. In [2] L. Kaloujnine gives a construction which shows that any p-group of class 2, where p is an odd prime, occurs as a circle group. In the other direction, this note obtains a restriction on the central series of p-groups which can occur as circle groups. A direct consequence of this result is that a group of order p^n which occurs as a circle group can have class at most $(n+1)/2$. An example is presented which shows that this bound is the best possible.

If S and T are subgroups of a group [subrings of a ring], with $T \subseteq S$, then $[S:T]$ will denote the index of T in S [the index of T in S considered as additive subgroups of the ring]. The letter p will always denote a prime.

THEOREM. *Let G be a p-group. If G occurs as the circle group of a nilpotent ring, then G has a central series*

$$G = Z_0 \supset Z_1 \supset \cdots \supset Z_c = 1$$

in which $[Z_{i-1}:Z_i] \geqq p^2$ *for* $1 \leqq i < c$.

REMARK 1. Note that the theorem makes no assertion concerning $[Z_{c-1}:Z_c] = |Z_{c-1}|$. To see that $|Z_{c-1}| = p$ can occur consider the ring of p elements in which all products are 0.

REMARK 2. To show that the bound on indices in this result cannot be improved, consider the following example. Let R be the ring generated by elements a, b with $p^n a = p^n b = 0$, $a^2 = ab = pa$, $ba = b^2 = pb$. The terms of the upper central series of the circle group of R coincide with the multiples of R of the form $p^k R$, $0 \leqq k \leqq n$, so all terms of the upper central series ascend by steps of p^2.

REMARK 3. Combining the above theorem with the construction of L. Kaloujnine [2] for $p \neq 2$, or with a direct construction for $p = 2$, one immediately obtains:

COROLLARY. *All groups of orders dividing p^3 occur as circle groups of nilpotent rings. A group of order p^4 occurs as a circle group if and only if it is Abelian or has class* 2.

To prove the theorem we shall first establish the following

LEMMA. *Let R be a nilpotent ring such that* $[R^{n-2}:R^{n-1}]$ *is a prime, where*

$R^n=0$, $R^{n-1}\neq 0$, *and* $n\geqq 3$. *Then* R^{n-2} *is in the center of* R.

Proof. Let $b=r_1r_2\cdots r_{n-2}$ be a product of length $n-2$ in R, with $b\notin R^{n-1}$. Since $[R^{n-2}:R^{n-1}]=p$, for some prime p, it follows that $pb\in R^{n-1}$, and b together with R^{n-1} spans R^{n-2}. Since R^{n-1} annihilates R, the center of R contains R^{n-1}. Assume R^{n-2} is not in the center. Then there is an element $a\in R$ with $ab\neq ba$. Hence either $ab\neq 0$ or $ba\neq 0$. To be definite suppose $ab\neq 0$.

We shall now show by a finite induction that $b\equiv xa^{n-2} \pmod{R^{n-1}}$ for some integer $x\not\equiv 0 \pmod p$. Suppose, for some integer k, $0\leqq k\leqq n-3$, that

$$b\equiv x_ka^kr_1r_2\cdots r_{n-k-2} \pmod{R^{n-1}} \qquad \text{for some integer } x_k\not\equiv 0 \pmod p. \tag{1}$$

Since R^{n-1} annihilates R, (1) implies $ab=(x_ka^{k+1}r_1\cdots r_{n-k-3})r_{n-k-2}$. Since $0\neq ab\in R^{n-1}$, it follows that $s=x_ka^{k+1}r_1\cdots r_{n-k-3}\notin R^{n-1}$. But $s\in R^{n-2}$, so $b\equiv ys \pmod{R^{n-1}}$ for some integer $y\not\equiv 0 \pmod p$. Setting $x_{k+1}=yx_k$, we obtain (1) with k replaced by $k+1$.

Condition (1) trivially holds for $k=0$, so by induction we obtain $b\equiv x_{n-2}a^{n-2} \pmod{R^{n-1}}$ for some integer $x_{n-2}\not\equiv 0 \pmod p$. Since, finally, R^{n-1} annihilates R, we have $ab=x_{n-2}a^{n-1}=ba$, which contradicts the assumption that $ab\neq ba$. Hence R^{n-2} is central in R.

Proof of Theorem. Let R be a nilpotent ring whose circle group G is a p-group. Since R is nilpotent, the powers of R form a finite descending central series for G. We shall select a subset,

$$R=Z_0\supset Z_1\supset\cdots\supset Z_c=0,$$

which forms a central series for G, and for which $[Z_{i-1}:Z_i]\geqq p^2$, $1\leqq i<c$. To do so, let us define $Z_0=R$ and proceed by induction. Suppose $Z_i=R^k\neq 0$. If $[R^k:R^{k+1}]\geqq p^2$, or if $R^{k+1}=0$, define $Z_{i+1}=R^{k+1}$. If $[R^k:R^{k+1}]=p$ and $R^{k+1}\neq 0$, then application of the lemma to the ring R/R^{k+2} shows that R^k/R^{k+2} is central in R/R^{k+2}. In this case define $Z_{i+1}=R^{k+2}$. This process terminates when $Z_i=0$, and the description of the desired central series is then complete.

Acknowledgment. The author wishes to thank D. T. Price of the University of Chicago for helpful comments and the United States Atomic Energy Commission for financial support.

References

1. N. Jacobson, The radical and semisimplicity for arbitrary rings, Amer. J. Math., 67 (1945) 300–320.

2. L. Kaloujnine, Zum Problem der Klassification der endlichen metabelschen p-Gruppen Wiss. Z. Humboldt-Univ. Berlin, Math.-Nat. Reihe, 4 (1955) 1–7.

A MEAN VALUE THEOREM

R. J. EASTON, Indiana State University, and S. G. WAYMENT, Utah State University

I. Introduction: Let μ be Lebesgue measure on E^n and let T be a finitely additive set function defined on the intervals R_n in E^n. A sequence $\{I_i\}$ of intervals in E^n is said to *converge* to $x\in E^n$, denoted by $I_i\to x$, provided $x\in I_i$ for

each i and $\lim_{i\to\infty} d(I_i)=0$, where $d(I)$ is the diameter of I. A sequence of intervals is said to be *regular* and α is the *parameter of regularity* provided that for each I_i there exists a cube $J_i \supset I_i$ so that $\mu(I_i)/\mu(J_i) \geqq \alpha$. The upper and lower regular derivatives of T are defined as

$$\overline{T'^*}(x) = \sup_{I_i\to x} \overline{\lim_{i}} \frac{T(I_i)}{\mu(I_i)} \quad \text{and} \quad \underline{T'^*}(x) = \inf_{I_i\to x} \underline{\lim_{i}} \frac{T(I_i)}{\mu(I_i)}$$

respectively, where the sup and inf are taken over regular sequences which converge to x. If $\overline{T}'^*(x)=\underline{T}'^*(x)$, then the *regular derivative* T'^* is said to exist [2]. If T'^* exists at each point p in an interval $R_n \subset E^n$, then L. Misik has shown [1] a mean value theorem holds.

We give here a simpler and more direct proof of his result using the additional hypothesis that T is absolutely continuous with respect to μ. For notational advantages we restrict our proof to E^2.

Let $R=[a, b; c, d]=\{(x, y): x\in(a, b),\ y\in(c, d)\}$. Denote the area of R by $A(R)$, and define $g_R(x, y)=T(R')$, where (x, y) is the mid-point of R' and R' has its edges parallel to the edges of R and is congruent to R. Since T is absolutely continuous with respect to μ, g_R is a continuous function of x and y.

II. The Mean Value Theorem. *If $T'^*(p)$ exists at each point p of a closed rectangle R and T is absolutely continuous with respect to μ, then there exists a point $q\in R$ such that $T'^*(q)=T(R)/A(R)$.*

Proof. Let $R_1=R=[a, b; c, d]$ and divide R_1 into four rectangles using the lines $x=a+h/2$, $y=c+k/2$ where $h=b-a$ and $k=d-c$. Denote the rectangles beginning in the lower left hand corner and proceeding counterclockwise by R_{11}, R_{12}, R_{13}, R_{14}, and observe that each of the four rectangles is similar to R_1. Since $\sum_{i=1}^{4} T(R_{1i})=T(R_1)$, there must exist a j_1 and a j_2 such that

$$T(R_{1j_1}) \leqq (1/4)T(R_1), \qquad T(R_{1j_2}) \geqq (1/4)T(R_1).$$

We consider the case $j_i=1$ and $j_2=3$. The other cases follow in a similar manner. Let $\alpha=k/h$ and define the function $f(t)=g_{R_{11}}(a+t+h/4,\ c+\alpha t+k/4)$. Then $f(0)=T(R_{11})$, $f(h/2)=T(R_{13})$, and f is a continuous function of t for $0\leqq t\leqq h/2$. Hence there exists a $t_0\in[0, h/2]$ such that $f(t_0)=(\frac{1}{4})T(R_1)$. The number t_0 determines a rectangle R_2. We proceed inductively to define a nested sequence of closed rectangles $\{R_i\}$, geometrically similar to R_1, and with sides parallel to R_1, such that

$$\text{(i)}\ T(R_{i+1}) = (\tfrac{1}{4})T(R_i) \qquad \text{and} \qquad \text{(ii)}\ A(R_{i+1}) = (\tfrac{1}{4})A(R_i).$$

By the nested interval theorem there exists exactly one point $q\in\bigcap_{i=1}^{\infty} R_i$, and

$$T'^*(q) = \lim T(R_i)/A(R_i) = T(R_1)/A(R_1).$$

REMARK 1. By changing the selection process slightly for R_i, $i=3, 4$ one can insure $q\in\mathrm{Int}(R)$.

REMARK 2. The derivative T'^* can be shown to have an interesting Darboux

property. Suppose G is an open set such that $T'^*(p)$ exists for each $p \in G$. Suppose p, $q \in G$ and $T'^*(p) = \alpha < T'^*(q) = \beta$ and let γ be between α and β. If $\overline{pq}$ is an arc in G and $\epsilon > 0$, then there exists a point x within ϵ of $\overline{pq}$ such that $T'^*(x) = \gamma$. This is easily seen to be true by obtaining two congruent rectangles R_p and R_q with diameter less than ϵ and centered at p and q respectively and such that $T(R_p)/\mu(R_p)$ and $T(R_q)/\mu(R_q)$ are respectively less than and greater than γ. By further restricting the size of R_p and R_q one can ensure that for each $w \in \overline{pq}$ there exists a rectangle R_w congruent to R_p and R_q such that $R_w \subset G$ and has edges parallel to the coordinate axes. Now since g_{R_p} is continuous on $\overline{pq}$ there exists a point $s \in \overline{pq}$ and a rectangle R_s centered at s and contained in G such that $T(R_s)/\mu(R_s) = \gamma$. Finally, an application of the mean value theorem yields the desired x.

REMARK 3. The requirement that T be absolutely continuous with respect to μ was needed only to insure that $g_{R'}$ be a continuous function of x and y for each $R' \subset R$. This leads one to the following conjecture:

CONJECTURE. *If T'^* exists for each $q \in R$, then $g_{R'}$ is a continuous function of x and y for each $R' \subset R$.*

If this conjecture were true, then the theorem given here is equivalent to Misik's theorem [1]. In fact if $g_{R'}$ could be shown to have an intermediate value property, then the theorems would be equivalent.

References

1. L. Misik, Der Mittelwertsatz für additive Intervalfunktionen, Func. Math., 45 (1957) 64–70.

2. M. E. Munroe, Introduction to Measure and Integration, Addison-Wesley, Reading, Mass., 1953.

A COMPACT NON-METRIZABLE SPACE SUCH THAT EVERY CLOSED SUBSET IS A G-DELTA

ROY A. JOHNSON, Washington State University

1. Introduction. By the *Borel sets* and *Baire sets* of a locally compact space we mean the σ-rings generated by compact sets and compact G_δ sets, respectively [1, pp. 219–220]. If X and Y are locally compact spaces, it is well known that the Baire sets of $X \times Y$ form precisely the product σ-ring of the Baire sets in X with the Baire sets in Y. See [1, 51.E]. It is therefore very tempting to believe that the Borel sets in $X \times Y$ form precisely the product σ-ring of the Borel sets in X with the Borel sets in Y, but unfortunately this is false. Worse yet, it is possible that the Borel sets and Baire sets of a space X may coincide (i.e., each compact set may be a G_δ), but that the Borel sets of $X \times X$ may form a strictly larger class than the Baire sets of $X \times X$. Such is the case for a compact non-metrizable space in which each closed subset is a G_δ. Let us see why.

We notice that if X is a compact Hausdorff space, then the Borel sets and

Baire sets of $X \times X$ coincide only if X is metrizable. For, if the diagonal D of $X \times X$ is a Baire set, then D is a compact G_δ. See [1, 51.D]. Hence, there exists a decreasing sequence of compact neighborhoods of D whose intersection is D. This sequence of decreasing neighborhoods is a base for the neighborhood system of D [2, Exercise 5.F(a)], and hence for the uniformity of X, so that X is metrizable [2, 6.13, 6.30].

2. The example. The space we construct can be viewed in three ways: as a subset of the reals with a suitable topology, as a subspace of a product of discrete spaces, or as a subspace of an ordered set having the order topology. We construct three homeomorphic spaces having the promised properties.

Let $X = [-1, 1) = \{x: -1 \leqq x < 1\}$. Let $\mathcal{S}$ be the family of subsets A such that A or $X - A$ has the form $[-b, b)$ for some b such that $0 \leqq b \leqq 1$. Then $\mathcal{S}$ is a subbase for a topology on X. If $\mathcal{B}$ is the base derived from $\mathcal{S}$, then

$$\mathcal{B} = \{[-b, -a) \cup [a, b): 0 \leqq a \leqq b \leqq 1\}.$$

Since members of $\mathcal{S}$ are both open and closed, so are base members.

THEOREM 2.1. *The space X is Hausdorff.*

Proof. Given two distinct points of X, then one has smaller absolute value than the other or the points are negatives of each other. In the first case, suppose $|a| < |b|$. Choosing c such that $|a| < c < |b|$, we have $a \in [-c, c)$ and $b \in X - [-c, c)$. In the second case, suppose $b = -a$, where $a > 0$. Then $b \in [-a, a)$ and $a \in X - [-a, a)$.

To construct the second example, let G be the class of all functions from $[0, 1]$ into the discrete space of two elements, and let G have the product topology. Then G is a compact Hausdorff space. The desired example is a subspace Y, defined as follows: Let Y be the class of all nonconstant increasing functions with the relative topology. That is,

$$Y = \{g \in G: g(0) = 0, g(1) = 1, \quad \text{and} \quad g(i) \leqq g(j) \text{ if } 0 \leqq i \leqq j \leqq 1\}.$$

Then Y consists of those functions which are 0 to the left of some point j, called a *jump point*, and 1 to the right of that point. We shall call a function g *lower jump* or *upper jump* at its jump point j, depending on whether $g(j)$ is 0 or 1. A function $g \in Y$ is clearly characterized by its jump point and the knowledge of whether it is lower jump or upper jump.

THEOREM 2.2. *The subspace Y is closed, and hence compact, in G.*

Proof. Suppose $f \in G - Y$. If $f(0) = 1$ [$f(1) = 0$], then $\{g \in G: g(0) = 1\}$ [resp., $\{g \in G: g(1) = 0\}$] is a neighborhood of f which is disjoint from Y. In the remaining case, there exist i and j such that $0 < i < j < 1$ and such that $f(i) > f(j)$. Necessarily, $f(i) = 1$ and $f(j) = 0$. Let

$$U = \{g \in G: g(i) = 1\} \cap \{g \in G: g(j) = 0\}.$$

Then U is a neighborhood of f which is disjoint from Y.

THEOREM 2.3. *The spaces X and Y are homeomorphic, and hence X is compact.*

Proof. Define S: $Y \to X$ as follows: If $g \in Y$ is lower jump at j, let $S(g) = j$. If $g \in Y$ is upper jump at j, let $S(g) = -j$. Then given $S(g)$, the jump point of g is $|S(g)|$, and g is lower or upper jump, depending on whether $S(g) \geqq 0$ or $S(g) < 0$. Thus, S is one-to-one, and it is clear that S is onto.

To show that S is continuous, we show that the inverses of subbase members of X are open sets in Y. See [2, 3.1]. If U is a subbase member of X having the form $[-b, b)$, then

$$S^{-1}[U] = \{g \in Y: g(b) = 1\},$$

an open set in Y. On the other hand, if $X - U = [-b, b)$, then

$$S^{-1}[U] = \{g \in Y: g(b) = 0\},$$

an open set in Y. Hence S is a continuous one-to-one mapping of the compact space Y onto the Hausdorff space X, which shows that S is a homeomorphism.

Incidentally, X could be shown directly to be compact by showing that every covering by subbase members can be reduced to a covering by two such members.

To construct the third example, let H be the unit square with dictionary order, denoted by $\prec$, and let H have the order topology [2, Exercise 5.J]. Then H is a compact Hausdorff space. The desired example is a subspace Z, defined as follows: Let

$$Z = \{(x, y) \in H: y = 0 \text{ or } 1\} - \{(0, 0), (1, 1)\}.$$

That is, Z consists of the lower and upper edges of H, without the first and last points of H. Let Z have the relative topology. Since Z is clearly closed in H, we see that Z is compact.

THEOREM 2.4. *The spaces X and Z are homeomorphic.*

Proof. Define $T: Z \to X$ by $T(x, 0) = -x$ and $T(x, 1) = x$. Clearly T is one-to-one and onto. We show that T is continuous by showing $T^{-1}[U]$ is open for each subbase member U of X. If U has the form $[-b, b)$, then

$$\begin{aligned} T^{-1}[U] &= \{(x, 0): -b \leqq -x < 0\} \cup \{(x, 1): 0 \leqq x < b\} \\ &= \{(x, y) \in Z: (x, y) \prec (b, 1)\}, \end{aligned}$$

an open set in Z. Similarly, if $X - U = [-b, b)$, then

$$T^{-1}[U] = \{(x, y) \in Z: (b, 0) \prec (x, y)\},$$

an open set in Z. Since T is a one-to-one continuous mapping of the compact space Z onto X, we see that T is a homeomorphism.

THEOREM 2.5. *The space X is not metrizable.*

Proof. Since X is compact, it would have a countable base if it were metrizable. But each $x \in [-1, 0)$ is a member of the open set $[x, -x)$, and each such pair of x and $[x, -x)$ would require a distinct member of the proposed countable base, which is impossible.

THEOREM 2.6. *The space X is a compact non-metrizable space such that each closed subset is a G_δ.*

Proof. We show that if U is open, then U is a countable union of base members and is therefore an F_σ. It follows that each closed subset is a G_δ, and this is all that needs to be shown.

For each $x \in U$, let V_x be a base member such that $x \in V_x \subset U$. Then $U = \cup \{V_x : x \in U\}$. We show that U can be covered by a countable subfamily of $\{V_x : x \in U\}$. Now by a *Euclidean interval* we shall mean an open interval in the usual sense. Let

$$W = \{y \in U : y \in W_y \subset V_x \text{ for some } x \in U \text{ and some Euclidean interval } W_y\}.$$

Clearly $W = \cup \{W_y : y \in W\}$. Then there exists a countable subfamily of $\{W_y : y \in W\}$, say W_n, such that $W = \cup W_n$. For each n, we choose $V_n \in \{V_x : x \in U\}$ such that $W_n \subset V_n$. Necessarily, $W \subset \cup V_n$. Hence, to show that U can be covered by a countable subfamily of $\{V_x : x \in U\}$, we need only show that $U - W$ is countable. To do this, we construct a one-to-one mapping from $U - W$ into the rational numbers. For each $x \in U - W$, choose a rational number r_x such that $x < r_x$ and such that $[x, r_x) \subset V_x$. The mapping $x \to r_x$ is one-to-one. Otherwise there would exist x and y in $U - W$ such that $x < y$ and such that $r_x = r_y$. But then $y \in (x, r_x) \subset V_x$, so that $y \in W$, a contradiction. Hence, $U - W$ is countable, and the proof is complete.

3. An observation. A Borel measure μ is called *completion regular* if each Borel set lies in the domain of the completion of the Baire contraction of μ [1, p. 230]. Equivalently, μ is completion regular if

$$\mu(E) = \sup \{\mu(C) : C \text{ is a compact } G_\delta \text{ and } C \subset E\}.$$

Letting X be the space in this note, we notice that each Borel measure on X is completion regular since the Borel sets and Baire sets coincide. But, interestingly, there exist Borel measures on $X \times X$ which fail to be completion regular.

For, let μ be any Borel measure on $X \times X$ such that the diagonal D has positive measure and such that each countable set has zero measure. (If $T(x, x) = x$ for each $x \in X$ and if m is Lebesgue measure on the line, then $\mu(E) = m(T[E \cap D])$ is such a measure.) Now if C is a compact subset in D, we let

$$A = \{(x, -x) : (x, x) \in C \text{ or } (-x, -x) \in C\}.$$

Then every open set containing C contains all but finitely many members of A, so that each G_δ containing C contains all but countably many members of A. Hence, if C is a compact G_δ, then C must be countable. Hence, μ fails to be completion regular.

References

1. P. R. Halmos, Measure Theory, Van Nostrand, Princeton, N. J., 1950.
2. J. L. Kelley, General Topology, Van Nostrand, Princeton, N. J., 1955.

A GROUP EPIMORPHISM IS SURJECTIVE

C. E. LINDERHOLM, Reading University

1. Introduction. The reader should be familiar with the most elementary part of a first course on group theory. Such a course may not contain the following definitions. We write $\phi: F\to G$ to mean that F and G are groups and ϕ is a homomorphism defined on F and taking values in G. If $\psi: G\to H$ the composite homomorphism from F to H is written $\phi\psi$, according to the usage of some algebraists. If $x\in F$ the image under ϕ is written $(x)\phi$. A homomorphism ψ that cancels on the right in every equation $\phi_1\psi=\phi_2\psi$ is a *monomorphism.* More precisely, if $\psi: G\to H$ then ψ is a monomorphism if for each group F and for each pair ϕ_1, ϕ_2 of homomorphisms from F to G such that $\phi_1\psi=\phi_2\psi$, the equation $\phi_1=\phi_2$ holds. A homomorphism ϕ that cancels on the left in every equation $\phi\psi_1=\phi\psi_2$ is an *epimorphism.* If $\phi: F\to G$ and if an inverse homomorphism ψ from G to F exists—i.e., one such that $(x)\phi\psi=x$ for each element x of F and $(y)\psi\phi=y$ for each element y of G—then ϕ is an *isomorphism.*

Suppose ϕ: $F\to G$. Then ϕ is a monomorphism if and only if ϕ is injective; an epimorphism if and only if surjective; and an isomorphism if and only if bijective.

(Injective, surjective, and bijective are the Norman terms for the Anglo-Saxon one-to-one, onto, one-to-one and onto.) The three assertions just made become six if they are analyzed into their "if" and "only if" parts. Of the six, five are very easy to prove; the hardest is the assertion of the title.

2. Proof. The present proof closely resembles one due to S. Eilenberg and J. Moore [3]; moreover, it seems to be well known to those who work with categories. A proof using the slightly sophisticated concept of free products with amalgamation is easy once one knows that concept. Such a proof, however, since it involves constructing a possibly infinite group, fails to yield the interesting corollary that *an epimorphism in the category of finite groups is surjective.*

Let ϕ be an epimorphism from a group G to a group H, and let A be the image subgroup. We must show that $A=H$. To do this we construct two homomorphisms ψ_1, ψ_2 from H to a group K such that $\phi\psi_1=\phi\psi_2$, and we use the resulting equation $\psi_1=\psi_2$ to prove that $A=H$.

Let H/A be the set of all right cosets of A; let A' be something that is not a right coset of A, and let S be the set $H/A\cup\{A'\}$. Let K be the group of all permutations (bijections) of S. If h, h_1, $h_2\in H$ and if Ah_1 and Ah_2 are the same coset, then $A(h_1h)$ and $A(h_2h)$ are the same coset, and hence the function from A/H to A/H that sends Ah' to $A(h'h)$ is well defined. It is easily seen to be bijective. (In fact its inverse is given by $Ah'\to Ah'h^{-1}$.) If A' is sent to itself, this

defines a bijection of S, which we write $(h)\psi_1$. The function ψ_1 from H to K so defined is easily seen to be a homomorphism. Let σ be the permutation of S that interchanges A and A' and leaves everything else fixed, and if $h \in H$ write $(h)\psi_2$ for the composite permutation $\sigma^{-1}((h)\psi_1)\sigma$. Then the function ψ_2 is the composition of ψ_1 with an inner automorphism of K, and hence is a homomorphism from H to K.

If $a \in A$, then $(a)\psi_1$ leaves both A and A' fixed, and σ leaves every other element of S fixed. Hence σ and $(a)\psi_1$ commute, and $(a)\psi_2 = \sigma^{-1}((a)\psi_1)\sigma = (a)\psi_1$. Since ψ_1 and ψ_2 agree on the range of ϕ, the equation $\phi\psi_1 = \phi\psi_2$ holds. Since ϕ is an epimorphism, $\psi_1 = \psi_2$. Hence $(h)\psi_1$ commutes with σ for each element h of H. Since $(h)\psi_1$ leaves A' fixed and σ interchanges A' and A, it follows that $(h)\psi_1$ leaves A fixed. Since $(h)\psi_1$ sends A to Ah, it follows that $h \in A$. Since h was an arbitrary element of the group H, it follows that $A = H$ and ϕ is surjective.

References

1. S. Lang, Algebraic Structures, Addison-Wesley, Reading, Mass., 1965.
2. B. Mitchell, Theory of Categories, Academic Press, New York, 1965.
3. S. Eilenberg and J. Moore, Foundations of Relative Homological Algebra, Memoirs of the AMS, 55 (1965).

RESEARCH PROBLEMS

EDITED BY VICTOR KLEE

In this Department the Monthly presents easily stated research problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Victor Klee, Department of Mathematics, University of Washington, Seattle, WA 98105.

DOES THERE EXIST A "FOUR NORMALS TRIANGLE"?

H. GUGGENHEIMER, Polytechnic Institute of Brooklyn

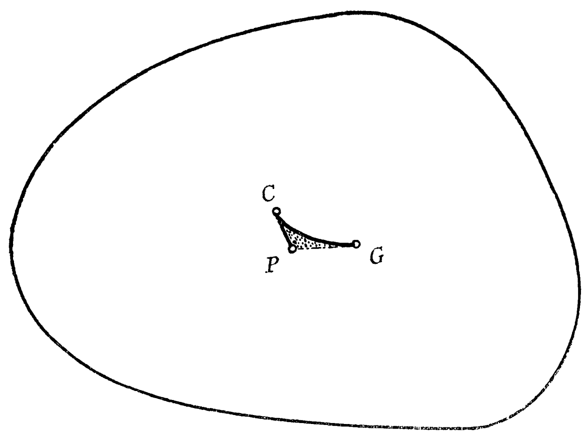
There are always at least two normals from an interior point of an oval (a closed, smooth, convex curve) to the curve: the segments of minimum and maximum length from the point to the oval. There exists quite a body of literature connected with the determination of points from which more than two normals can be drawn to a curve. It is easy to see that for any oval that is not a circle there exists a domain of positive area all of whose points are "four-normals points," i.e., from which at least four normals can be drawn to the oval. The only four-normals point of the circle is its center.

Of particular interest in this respect are three centers of mass distributions connected with an oval: the centroid G, the perimeter centroid P (center of mass for a density 1 on the boundary curve), and the curvature centroid or Steiner point C (center of mass for a density proportional to the radius of curvature on the boundary curve). Hayashi [1] has shown that C is a four-normals

point, and Bose [2] did the same for G. These proofs use Fourier analysis or some equivalent tool [3]. Chakerian and Stein [4] proved that P is a four-normals point; their proof is a simple continuity argument. Bose and Roy [5] noted that the normals to a curve are identical with the normals to the outer parallel curves for all positive distances (the boundary of the union of all circles of a constant radius and centers in or on the oval). If $G(c)$, $P(c)$, $C(c)$ are the corresponding points for the parallel curve in distance $c \geqq 0$ and A, L, are, respectively, surface area and perimeter of the given oval, we have by formulas of Nicliborc [6]

$$\begin{aligned} G(c) &= (A + Lc + \pi c^2)^{-1}\{AG + cLP + c^2\pi C\}, \\ P(c) &= (L + 2\pi c)^{-1}\{LP + 2\pi c C\}, \\ C(c) &= C. \end{aligned}$$

Hence, $C = \lim_{c\to\infty} G(c) = \lim_{c\to\infty} P(c)$. The points $G(c)$ are on a parabolic arc joining G and C; the points $P(c)$ fill a segment CP of the tangent to the arc at C; and the endpoint P is the intersection of that tangent with the tangent to the parabola at G. In this way, one obtains a "triangle" CPG formed by the arc of parabola GC and its tangents at the endpoints. It is known, then, that all points of the arc CG and the segment CP are four-normals points. Question: Is it true that all points of CPG and its interior are four-normals points? The known methods do not seem to work, probably because the methods needed for CG and for CP have nothing in common.



References

1. T. Hayashi, Some geometrical applications of Fourier series, Rend. Circ. Mat. Palermo, 1 (1926) 96–102.

2. R. C. Bose, A note on the convex oval, Bull. Calcutta Math. Soc., 27 (1935) 54–60.

3. H. Guggenheimer, Geometrical applications of integral calculus, *Lectures on Calculus*, K. O. May, ed., Holden-Day, San Francisco 1967, pp. 75–93.

4. G. D. Chakerian and S. K. Stein, On the centroid of a homogeneous wire, Mich. Math. J., 11 (1964) 189–192.

5. R. C. Bose and S. N. Roy, Some properties of the convex oval with reference to its curvature centroid, Bull. Calcutta Math. Soc., 27 (1935) 78–86.

6. W. Nicliborc, Über die Lage des Schwerpunktes eines ebenen konvexen Bereiches und die Extrema des logarithmischen Flächenpotentials eines konvexen Bereiches, Math. Z., 36 (1932) 161–165.

CLASSROOM NOTES

EDITED BY DAVID DRASIN

Manuscripts for this Department should be sent to David Drasin, Division of Mathematical Sciences, Purdue University, Lafayette, IN 47907.

SATURATION OF MEASURES

W. G. DOTSON, JR., North Carolina State University

Suppose (X, B, μ) is a measure space. A set $E\subset X$ is said to be *locally measurable* provided that if $F\in B$ and $\mu F<\infty$ then $F\cap E\in B$. The measure μ is said to be *saturated* [1, p. 221] provided every locally measurable set is measurable. It is well known that a measure μ induced by an outer measure μ^* is complete. It has recently been suggested [1, p. 253, Problem 2] that a measure μ induced by an outer measure μ^* must also be saturated. That this is not true can be seen by the following example:

Let X be any uncountable set, and define μ^* by $\mu^*\varnothing=0$, $\mu^*E=1$ if E is any nonempty countable subset of X, and $\mu^*E=\infty$ if E is any uncountable subset of X. It is easily seen that μ^* is an outer measure and that the only μ^*-measurable sets are $\varnothing$ and X. Indeed, if $E\subset X$ and $\varnothing\neq E\neq X$ then letting $A=\{p, q\}$ where $p\in E$, $q\in X\backslash E$, we have $\mu^*A=1<1+1=\mu^*(A\cap E)+\mu^*(A\backslash E)$. So the measure μ induced by μ^* is given by $\mu=\mu^*|B$ where $B=\{\varnothing, X\}$. It is clear that each set $E\subset X$ is locally measurable, since the only $F\in B$ with $\mu F<\infty$ is $F=\varnothing$. Thus μ is not saturated.

On the other hand, it is well known that if μ_0 is a measure on an algebra B_0, then the outer measure μ^* induced by μ_0 is *regular*, e.g., see [1, p. 260]. The outer measure μ^* then induces a measure μ on a σ-algebra B containing B_0, and $\mu|B_0=\mu_0$. It is in this context of the Carathéodory extension theorem that measures μ induced by outer measures μ^* are most frequently encountered.

THEOREM. *If μ^* is a regular outer measure, then the measure μ induced by μ^* is saturated.*

Proof. Let B denote the σ-algebra of all μ^*-measurable subsets of X, so that $\mu=\mu^*|B$. Suppose E is a locally measurable subset of X. For any $A\subset X$ we must show $\mu^*A\geqq\mu^*(A\cap E)+\mu^*(A\backslash E)$. Since this is obviously true if $\mu^*A=\infty$, we

suppose $\mu^*A<\infty$. Since μ^* is regular there exists $C\in B$ such that $C\supset A$ and $\mu C=\mu^*C=\mu^*A<\infty$. It now follows at once from the local measurability of E that $C\cap E\in B$, and so $C\setminus E=C\setminus(C\cap E)\in B$ also. Hence

$$\begin{aligned}\mu^*A &= \mu C = \mu(C\cap E)+\mu(C\setminus E)\\ &= \mu^*(C\cap E)+\mu^*(C\setminus E)\\ &\geqq \mu^*(A\cap E)+\mu^*(A\setminus E),\end{aligned}$$

since $C\supset A$ and μ^* is monotone. Q.E.D.

COROLLARY 1. *If the outer measure μ^* is induced by a measure on an algebra, then the measure μ induced by μ^* is saturated.*

COROLLARY 2. *The outer measure in the above example is not regular, and is therefore not induced by any measure on an algebra.*

Reference

1. H. L. Royden, Real Analysis, 2nd ed., Macmillan, New York, 1968.

AN INEQUALITY ABOUT COMPLEX NUMBERS

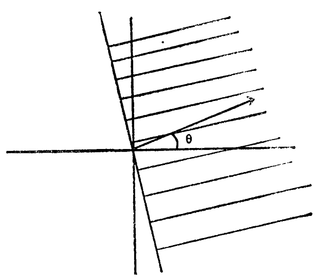
W. W. BLEDSOE, University of Texas

In this note we give a simple proof of the theorem listed below about complex numbers. This is a slightly stronger result than that given by Rudin (Walter Rudin, *Real and Complex Analysis*, Lemma 6.3, p. 119). His constant of 1/6 has been replaced by the best value of $1/\pi$.

THEOREM. *Every finite set $\{z_1, z_2, \cdots, z_n\}$ of complex numbers has a subset S such that*

$$\left|\sum_{z\in S_\theta} z\right| \geqq \frac{1}{\pi}\sum_{j=1}^{n} |z_j|.$$

Proof. For each θ, $0\leqq\theta<2\pi$, let S_θ be the set of those z_j such that $-\pi/2\leqq\theta-\arg(z_j)\leqq\pi/2$, and let $f(\theta)=\left|\sum_{z\in S_\theta} z\right|$.



We complete the proof by showing that

$$f(\theta') \geqq \frac{1}{\pi} \sum_{j=1}^{n} | z_j |$$

for some θ'.

Note that $f(\theta)$ is piecewise constant in θ,

$$\begin{aligned}
\left| \sum_{z \in S_\theta} z \right| &= \left| e^{-i\theta} \sum_{z \in S_\theta} z \right| = \left| \sum_{z \in S_\theta} z e^{-i\theta} \right| \\
&\geqq \left| \mathcal{R}\left(\sum_{z \in S_\theta} z e^{-i\theta} \right) \right| = \left| \sum_{z \in S_\theta} \mathcal{R}(z e^{-i\theta}) \right| \\
&= \left| \sum_{z \in S_\theta} | z | \cos(\theta - \arg z) \right| \\
&= \sum_{z \in S_\theta} | z | \cos(\theta - \arg z),
\end{aligned}$$

and hence

$$\begin{aligned}
\int_0^{2\pi} f(\theta) d\theta &= \int_0^{2\pi} \left| \sum_{z \in S_\theta} z \right| d\theta \\
&\geqq \int_0^{2\pi} \sum_{z \in S_\theta} | z | \cos(\theta - \arg(z)) d\theta \\
&= \sum_{j=1}^{n} \int_{\arg(z_j) - \pi/2}^{\arg(z_j) + \pi/2} | z_j | \cos(\theta - \arg(z_j)) d\theta \\
&= \sum_{j=1}^{n} 2 | z_j | .
\end{aligned}$$

Thus, by the extreme value theorem, there is a θ' between 0 and 2π for which

$$f(\theta') \geqq \frac{1}{2\pi} \sum_{j=1}^{n} 2 | z_j | .$$

The desired conclusion is at hand.

The constant $1/\pi$ is the best (largest) possible, as can be seen as follows. Let C_0 be the supremum of numbers C for which

$$\left| \sum_{z \in S} z \right| \geqq C \sum_{z \in T} | z | ,$$

where T is any nonempty, finite set of complex numbers and S is any subset of T. The above theorem shows that $C_0 \geqq 1/\pi$.

Now consider $z_j = \exp[i(2\pi/n)j]$, $j = 1, \cdots, n$. For this choice of points we

have, for each θ,

$$\lim_{n\to\infty}\left(\left|\sum_{z\in S_\theta} z\right| \Big/ \sum_{j=1}^{n} |z_j|\right) = \lim_{n\to\infty}\frac{1}{n}\left|\sum_{z\in S_\theta} z\right|$$

$$= \lim_{n\to\infty}\frac{1}{2\pi}\left|\frac{\pi}{n/2}\sum_{z\in S_\theta} z\right|$$

$$= \frac{1}{2\pi}\left|\int_{\theta-\pi/2}^{\theta+\pi/2} e^{i\phi}d\phi\right| = \frac{1}{2\pi}\cdot 2 = \frac{1}{\pi}.$$

Thus $C_0 \leqq 1/\pi$ and $C_0 = 1/\pi$, which shows that $1/\pi$ is the best such constant.

This technique readily lends itself to higher dimensions. If σ is the rotation-invariant measure on the unit sphere $S^{k-1} \subset R^k$, so normalized that $\sigma(S^{k-1}) = 1$, and

$$(1) \qquad c(k) = \int_{S_u} (q\cdot u)\sigma\, dq,$$

where u is any point on S^{k-1}, $q\cdot u$ is the usual scalar product in R^k, and S_u is the half-sphere

$$S_u = \{q \in S^{k-1}\colon q\cdot u \geqq 0\},$$

then we obtain the following:

THEOREM. *Every finite set* $\{p_1, p_2, \cdots, p_n\} \subset R^k$ *has a subset* E *such that*

$$\left|\sum_{p\in E} p\right| \geqq c(k)\sum_{j=1}^{n} |p_j|.$$

The $c(k)$, defined by (1), are the best such constants. Computation shows that $c(1) = 1/2$, $c(2) = 1/\pi$, $c(3) = 1/4$.

After this paper was accepted for publication the author became aware of an earlier paper, (Kaufman and Rickert, "An inequality concerning measures," Bul. AMS, 72 (1966), 672–676), which generalizes most of the results given in this paper. The treatment here is different and, hopefully, more readable.

RIEMANN INTEGRATION OF LIMIT FUNCTIONS

H. KESTELMAN, University College, London

1. The tendency to down-grade Riemann integration into 'calculus' gets some of its impetus from the fact that the more useful rules for integrating limit functions are hard to prove for Riemann integrals without introducing Lebesgue measure. It is frustrating to be confined to *uniform* convergence when we know from Lebesgue's theorem on dominated convergence that

$$(1) \qquad (R)\int_0^1 \lim_{n\to\infty} f_n(x)dx = \lim_{n\to\infty} (R)\int_0^1 f_n(x)dx,$$

whenever the f_n are Riemann-integrable and converge boundedly to a Riemann-integrable function.

The object of this note is to prove (1), in the conditions stated, without the apparatus of Lebesgue measure: this is done in Theorem 1. Applications to infinite and repeated integrals, and to differentiation under the integral sign, are also discussed; those who continue to preach Riemann integration may find the exercise interesting.

A referee, to whom I am grateful, points out that W. F. Osgood gave a discussion of the problem many years ago. It can be found in Amer. J. Math., 19 (1897) 155–190, with the title "Non-uniform convergence and the integration of series term by term." Osgood restricts himself to continuous functions, not necessarily uniformly bounded. His paper is interesting as a pioneer effort in set theory, but it could hardly be used for undergraduate instruction.

2. We take for granted in R_1 that

(a) a decreasing sequence of bounded closed nonempty sets has a nonempty intersection,

(b) every bounded open set is, in a unique way, the union of countably many disjoint open intervals.

If I is the interval (a, b) we write $|I|$ for $b-a$, and if $\{I_n\}$ are the constituent intervals of a bounded open set S we write $|S|$ for $\sum_n |I_n|$ and refer to it as the 'measure' of S. If T is then an open subset of S it is easy to show that

$$(2) \qquad |S| \leqq |T| + |U| \qquad \text{if } U \text{ is any open cover of } S - T.$$

This is all the 'measure theory' we use.

3. We shall need the following

LEMMA: *Let* $\{E_n\}$ *be a decreasing sequence of open subsets of* $(0, 1)$ *with* $\inf_n |E_n| > 0$; *then* $\bigcap_r E_n \neq \emptyset$.

Proof. What we have to prove is the weakest possible dilution of Lebesgue's theorem that $m(\overline{\lim}_n E_n) \geqq \lim(mE_n)$, but without using measure theory.

By (a) it is enough to find a decreasing sequence of open sets $\{U_n\}$ with $E_n \supset \overline{U}_n \neq \emptyset$. The idea is to select from each E_n a finite set of intervals; these are then 'shrunk' to provide a closed subset of E_n. The details ensuring that the closed sets decrease are a little tiresome, but the method is intuitive.

We choose a number θ, $0<\theta<1$, to be further restricted presently. If $I = (a-\delta, a+\delta)$ we write I^* for the 'shrunk' interval $(a-\theta\delta, a+\theta\delta)$. Clearly $\overline{I^*} \subset I$ and $I - I^*$ is covered by an open set with measure less than $2(1-\theta)|I|$.

Set $p = \inf_n |E_n|$. $\{I_n\}$ being the intervals of E_1, choose N so that $|I_{N+1}| + |I_{N+2}| + \cdots < \frac{1}{4}p$ and define $U_1 = I_1^* \cup I_2^* \cup \cdots \cup I_N^*$. Then $E_1 \supset \overline{U}_1 \neq \emptyset$ and $E_1 - U_1$ is covered by an open set with measure less than $\frac{1}{4}p + 2(1-\theta)|E_1|$. For each $r \geqq 2$, $E_r \cap U_1$ is open; since $E_r \cap cU_1 \subset E_1 - U_1$, it follows from (2) that

$$|E_r \cap U_1| \geqq |E_r| - (\tfrac{1}{4}p + 2(1-\theta)|E_1|) > \tfrac{1}{2}p$$

provided $2(1-\theta)|E_1|<\frac{1}{4}p$. If θ is so chosen then $\inf_{r\geqq 2}|E_r\cap U_1|>0$, and we may repeat the argument above on the decreasing sequence of open sets $E_r\cap U_1(r\geqq 2)$ to define an open set U_2 with $\emptyset\neq\overline{U}_2\subset E_2\cap U_1$ and $\inf_{r\geqq 3}|E_r\cap U_2|>0$. The process can be continued inductively to define a sequence $\{U_n\}$ as required.

4. We come now to the main

THEOREM 1. *Suppose that $f_0, f_1, f_2, \cdots$ are integrable-R over $[0, 1]$ and that $\{f_n(x)\}$ converges boundedly to $f_0(x)$ throughout $[0, 1]$. Then (1) is valid.*

Proof. We have only to prove that $\lim_{n\to\infty}(R)\int_0^1\{f_n(x)-f_0(x)\}dx=0$, and so we may without loss of generality assume that $f_0=0$. Now assume the theorem false.

It follows (multiply the f_n by a suitable constant) that a sequence $\{g_n\}$ exists with the following properties:

$$(\alpha)\qquad -1\leqq g_n(x)\leqq 1\quad\text{for all } n \text{ and all } 0\leqq x\leqq 1,\quad\text{and}\quad \lim_{n\to\infty} g_n(x)=0,$$

$$(\beta)\qquad \inf_n(R)\int_0^1 g_n(x)dx=p>0.$$

If we show that there is, for each n, an open set A_n in which $g_n(x)\geqq\frac{1}{2}p$, and $\inf_n|A_n|>0$, we may deduce from the lemma, on setting $E_n=A_n\cup A_{n+1}\cup\cdots$, that for some ξ in $[0, 1]$, $g_n(\xi)\geqq\frac{1}{2}p$ for infinitely many n. This contradicts $\lim_{n\to\infty}g_n(\xi)=0$ and so establishes the theorem.

To define A_n, let s_n be a lower Riemann sum of g_n with $s_n>\frac{3}{4}p$; let q_n be the sum of the lengths of those cells in which $\inf g_n\geqq\frac{1}{2}p$; the contribution to s_n from these cells is $\leqq q_n$ (by (α)), and that from the other cells is $\leqq\frac{1}{2}p(1-q_n)$. This means $\frac{3}{4}p\leqq s_n\leqq\frac{1}{2}p+q_n(1-\frac{1}{2}p)$, and so $q_n(1-\frac{1}{2}p)\geqq\frac{1}{4}p$. Thus, if A_n is the union of the interiors of those cells throughout which $g_n(x)\geqq\frac{1}{2}p$, we have $(1-\frac{1}{2}p)|A_n|\geqq\frac{1}{4}p$, and this completes the proof.

5. Theorem 1 may be extended to Cauchy-Riemann integrals. If $\lim_{X\to\infty}(R)\int_0^X f(x)dx$ exists and is finite we denote it by $(\mathrm{CR})\int_0^\infty f(x)dx$.

THEOREM 2. *Let $f_0, f_1, \cdots$ be nonnegative and defined on $[0, \infty)$ so that*

$$f_0(x)=\sum_{n=1}^{\infty}f_n(x)\quad\textit{and}\quad (R)\int_0^X f_n(x)dx$$

exists for all $0<X<\infty$ and $n\geqq 0$. Then

$$(3)\qquad (\mathrm{CR})\int_0^\infty f_0(x)dx=\sum_{n=1}^{\infty}(\mathrm{CR})\int_0^\infty f_n(x)dx$$

if the right-hand side is known to be finite or if the left-hand side is known to exist.

Proof. Choose X, $0<X<\infty$; then for $0\leqq x\leqq X$ and $N\geqq 1$, $\sum_{n=1}^{N} f_n(x) \leqq \sup_{0\leqq x\leqq X} f_0(x)$ and by Theorem 1

$$(R)\int_0^X f_0(x)dx = \sum_{n=1}^{\infty} (R)\int_0^X f_n(x)dx.$$

If R.H.S. of (3) is finite then

$$(R)\int_0^X f_0(x)dx \leqq \sum_{n=1}^{\infty} (\mathrm{CR})\int_0^{\infty} f_n(x)dx \qquad \text{for all } 0 < X < \infty.$$

Since $\int_0^X f_0(x)dx$ increases with X, this implies that L.H.S.$\leqq$R.H.S. in (3), but

$$(\mathrm{CR})\int_0^{\infty} f_0(x)dx \geqq \sum_{n=1}^{N} (\mathrm{CR})\int_0^{\infty} f_n(x)dx \qquad (N = 1, 2, \cdots),$$

which justifies (3).

If L.H.S. of (3) exists, then $(\mathrm{CR})\int_0^\infty f_n(x)dx$ exists for all n and

$$\sum_{n=1}^{N} (\mathrm{CR})\int_0^{\infty} f_n(x)dx \leqq (\mathrm{CR})\int_0^{\infty} f_0(x)dx \qquad (N = 1, 2, \cdots);$$

this implies that R.H.S. of (3) is finite and again (3) is justified.

When the functions are not of constant sign we have

THEOREM 3. *Let $f_0, f_1, \cdots$ be defined on $[0, \infty)$ so that*
(i) $f_0(x) = \sum_{n=1}^{\infty} f_n(x)$,
(ii) $(R)\int_0^X f_n(x)dx$ *exists for all* $0<X<\infty$ *and* $n\geqq 0$, *and*
(iii) $(\mathrm{CR})\int_0^\infty \sum_{n=1}^{\infty} |f_n(x)|dx$ *exists.*
Then (3) *is valid.*

Proof. For $0<X<X'<\infty$ and $r\geqq 0$ we have (using (iii))

$$\left|(R)\int_X^{X'} f_r(x)dx\right| \leqq (R)\int_X^{X'} \sum_{n=1}^{\infty} |f_n(x)|\,dx \to 0 \qquad \text{as } X\to\infty;$$

this means that $(\mathrm{CR})\int_0^\infty f_r(x)dx$ exists for all $r\geqq 0$, and so by (iii) and Theorem 2, both sides of (3) exist. For any real X the partial sums of $\sum_{n=1}^{\infty} f_n(x)$ are dominated in $[0, X]$ by $\sup_{0\leqq x\leqq X}\sum_{n=1}^{\infty} |f_n(x)|$, and so it follows from Theorem 1 that

$$(R)\int_0^X f_0(x)dx = \sum_{n=1}^{\infty} (R)\int_0^X f_n(x)dx$$
$$= \sum_{n=1}^{\infty}\left\{(\mathrm{CR})\int_0^{\infty} f_n(x)dx - (\mathrm{CR})\int_X^{\infty} f_n(x)dx\right\},$$

but

$$\left| \sum_{n=1}^{\infty} (\mathrm{CR}) \int_X^{\infty} f_n(x)dx \right| \leqq \sum_{n=1}^{\infty} (\mathrm{CR}) \int_X^{\infty} | f_n(x) | \, dx$$

$$= (\mathrm{CR}) \int_X^{\infty} \sum_{n=1}^{\infty} | f_n(x) | \, dx$$

(using Theorem 2), and the last integral$\to 0$ as $X\to\infty$ by (iii). This proves (3).

6. As an application to infinite repeated integrals we prove

THEOREM 4. *Let $f(x, y)$ be continuous and $f(x, y) \geqq 0$ when $x, y \geqq 0$. Then*

$$\int_0^{\infty} dy \int_0^{\infty} f(x, y)dx = \int_0^{\infty} dx \int_0^{\infty} f(x, y)dy, \tag{4}$$

assuming all the integrals exist in the Cauchy-Riemann sense.

Proof. If X and Y are natural numbers we have

$$\int_0^{Y} dy \int_0^{X} f(x, y)dx = \int_0^{X} dx \int_0^{Y} f(x, y)dy.$$

For $0 \leqq y \leqq Y$, $\int_0^X f(x, y)dx$ is dominated by $\sup_{0 \leqq y \leqq Y} \int_0^{\infty} f(x, y)dx$. So, by (5) and Theorem 1 (X being integral)

$$\int_0^{Y} dy \lim_{X\to\infty} \int_0^{X} f(x, y)dx = \lim_{X\to\infty} \int_0^{X} dx \int_0^{Y} f(x, y)dy;$$

hence

$$\int_0^{Y} dy \int_0^{\infty} f(x, y)dx \leqq \int_0^{\infty} dx \int_0^{\infty} f(x, y)dy,$$

which implies that L.H.S.$\leqq$R.H.S. in (4). The assumptions being symmetric in x and y, this proves (4).

Finally, the theorem on differentiation under the integral sign.

THEOREM 5. *Suppose $g(x, y)$ is defined in I, the finite interval of R_2 given by $a < x < b$, $|y - y_0| < \delta$, so that*

(i) *$(\partial/\partial y)g(x, y)$ is bounded in I and $(R)\int_a^b g_y(x, y_0)dx$ exists,*

(ii) *$(R)\int_a^b g(x, y)dx$ exists for $|y - y_0| < \delta$. Then for $y = y_0$*

$$\frac{d}{dy}\int_a^b g(x, y)dx = \int_a^b g_y(x, y_0)dx.$$

Proof by (i).

$$\frac{g(x, y_0 + h) - g(x, y_0)}{h} - g_y(x, y_0)$$

is bounded for $a<x<b$ and $0<|h|<\delta$, and for each x this function tends to zero with h. Hence, by the analogue of Theorem 1 for continuous parameters, we have, by (ii)

$$\lim_{h\to 0}\int_a^b\left\{\frac{g(x, y_0+h)-g(x, y_0)}{h}-g_y(x, y_0)\right\}dx=0$$

which, by (i), gives the result required.

MATHEMATICAL EDUCATION

EDITED BY J. G. HARVEY AND M. W. POWNALL

Material for this Department should be sent to either of the editors: J. G. Harvey, Department of Mathematics, University of Wisconsin, Madison, WI 53706; M. W. Pownall, Department of Mathematics, Colgate University, Hamilton, NY 13346.

AN INTRODUCTION TO DIFFERENTIAL CALCULUS

D. G. HERR, Duke University

How do you begin a first course in calculus on a note of excitement and anticipation? How do you convey the essence of what the course is about in two or three introductory lectures? These are some of the questions which faced two of us this spring in planning a revamped calculus sequence. We decided to give three introductory lectures: 1. An Introduction to Differential Calculus, 2. An Introduction to Integral Calculus, and 3. An Introduction to the Real Numbers. The order of the first two is rather arbitrary, but we felt that since a lecture on the real numbers was potentially the driest of the three we would build up some interest in the properties of the real numbers (especially inequalities) by having the introduction to calculus precede the lecture on real numbers. So far our ideas paralleled those in several texts.

The next problem we had to face was the content of the three lectures. Taking the last first, we decided on an axiomatic treatment of the real numbers with some historical notes such as the discovery of the irrationality of π in 1761 by Lambert [3] and the solution of the problem of "squaring the circle" as a result of C. L. F. Lindemann's proving in 1882 that π is not algebraic [3]. The lecture on the integral was not difficult to construct in view of T. M. Apostol's excellent introduction to integration in his calculus text [1]. However, the lecture on differential calculus was a problem.

The usual introduction to differential calculus makes use of physical concepts such as velocity and thus immediately the focus of the discussion is on the limit of the difference quotient. This seems to us unfortunate for two reasons. First this attention to the derivative does not generalize satisfactorily to functions of several variables in that the existence of partial derivatives is not a generaliza-

tion of the existence of a derivative. In addition this immediate obsession with the difference quotient obscures the actual problem being attacked and as a consequence such concepts as differentials often remain a mystery.

We take the position that the principal concern of differential calculus is the linearizing of functions, or more completely the identification of linearizable (differentiable) functions and their subsequent linearization. To be more explicit suppose f is a real valued function defined on the real line, R. In geometric terms linearizing f will mean finding a straight line which best approximates f. A student can readily visualize such a linearization and furthermore simple examples quickly demonstrate the futility of trying to obtain one straight line to approximate f over all of R in any reasonable way. In this way the point that we are concerned with local properties of functions in differential calculus is clearly and forcefully made. The problem now is to move from the geometric argument to an analytic one which will be explicit enough to answer two questions: "Which functions are linearizable at a point?" and "If a function is linearizable, what is the 'best' straight line approximation?" To effect this transition we ask the student to help formulate a definition of "best straight line approximation."

Since a nonvertical straight line in the plane has the general equation $y=ax+b$ and functions of the form $A=\{(x, y);\ y=ax+b\}$ are called affine functions, we want the "best affine approximation to f at a point."

It is not difficult to get the students to agree that one condition that a best affine approximation to f at x_0 should satisfy is

$$\text{(I)} \qquad A(x_0) = f(x_0),$$

where A is the approximating affine function. This establishes that $b=f(x_0)-ax_0$ so that we need only specify a, the slope of the line, to have A completely determined. A second reasonable condition to impose on A is that

$$\text{(II)} \qquad |A(x) - f(x)| \to 0 \quad \text{as } |x - x_0| \to 0,$$

i.e., the error made by approximating f by A at x becomes arbitrarily small as x tends to x_0. Examples easily demonstrate that (II) may pose no restriction whatever on A and is thus not a strong enough condition for our purposes. However, although the error $|A(x)-f(x)|$ gets small for any A satisfying (I) (f continuous at x_0) as $x\to x_0$, examples are again enough to demonstrate that the rate at which the error gets small is not the same for all A. If the relative error $|A(x)-f(x)|/|x-x_0|$ is used as a measure of the rate at which the error tends to zero, it is natural enough to require the best approximation to have the best rate, i.e.,

$$\text{(III)} \qquad |A(x) - f(x)|/|x - x_0| \to 0 \quad \text{as } |x - x_0| \to 0.$$

Note here that (III) implies (II). By noting that if A satisfies (I), $|A(x)-f(x)| = |f(x)-f(x_0)-a(x-x_0)|$, the student can readily see that (III) is equivalent to

$$\text{(IV)} \qquad \left|\frac{f(x) - f(x_0)}{x - x_0} - a\right| \to 0 \qquad \text{as } x \to x_0.$$

It can then be pointed out that if the "best affine approximation to f at x_0" is defined to be any A which satisfies (I) and (III), then the slope of the best fitting straight line, a, satisfies (IV) which means that a and thus A are uniquely determined. Furthermore (IV) then provides an answer to both original questions, i.e., the functions f which are linearizable at x_0 are precisely those for which the limit as $x \to x_0$ of $[f(x)-f(x_0)]/(x-x_0)$ exists as a real number, say $f'(x_0)$, and the best affine approximation is $A = \{(x, y): y = f'(x_0)(x-x_0)+f(x_0)\}$. Thus the best affine approximation to f at x_0 is defined to be any A which satisfies (I) and (III).

With this foundation the student can see limits of difference quotients, derivatives, and differentiability in perspective. Also the differential of f at x_0 as the linear part of the best affine approximation to f at x_0 is a straightforward concept without any of the mystery usually associated with a differential. Finally these ideas generalize directly to vector-valued functions of several variables with the understanding that linearizability and differentiability are synonymous [2].

References

1. T. M. Apostol, Calculus, Volume 1, Blaisdell, Waltham, Mass., 1961.
2. R. H. Crowell and R. E. Williamson, Calculus of Vector Functions, Prentice-Hall, Englewood Cliffs, N. J., 1962, 196–207.
3. G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, 4th ed., Oxford, London, 1960; 47, 173.

GENESIS*

Chapter 1

In the beginning Euclid created space.

2. And space was without structure, and void; and darkness was upon the face of deep space. And the spirit of Euclid moved.

3. And Euclid said let there be lines: and there were lines.

4. And Euclid saw the lines, that they were good: and Euclid divided the lines with points.

5. And Euclid called these lines rays. And the extent of the line was the first axiom.

6. And Euclid said let there be planes, and let them divide the points from the points.

7. And Euclid made planes and they divided the points which were under the planes from the points which were above the planes: and it was so.

8. And Euclid called the planes flat. And the evening and the morning were the second axiom.

9. And Euclid said let the points be gathered together unto one place, and let solids appear: and it was so.

10. And Euclid called the solids real; and the gathering together of the points he called space; and Euclid saw that it was good.

---

* This is a translation of an MS recently found in an ancient wine vessel in a cave in the mountains of Iraq. There are missing sections because of the fragmentary nature of the MS.

11. And Euclid said, let the space bring forth segments, the segments yielding fruit after their kind whose seed is in itself: and it was so.
12. And space brought forth segments, yielding segments upon segments to make number lines: and Euclid saw that it was good.
13. And the evening and the morning were the third axiom.
14. And Euclid said let there be orbs in the firmament of space to divide the inside from the outside and let them be signs for the seasons, and for days, and for years.
15. And let them be for gems in the firmament of space to give delight upon the earth: and it was so.
16, 17, 18. And Euclid made two great orbs the greater orb to rule the firmament and the lesser orb to rule the planes: he made the triangles also: and Euclid saw that it was good.
19. And the evening and the morning were the fourth axiom.
20, 21. And Euclid said let the points bring forth abundantly parallels so that alternate interiors are forever equal. And Euclid created great cubes, and squares, and every parallelogram that cometh, which the points brought forth abundantly, after their kind one upon another: and Euclid saw that it was good.
22. And Euclid blessed them, saying, be fruitful and multiply and with parallelograms and parallelepipeds fill the planes and space.
23. And the evening and the morning were the fifth axiom.
26. And Euclid said let us make a Mathematician in our image, after our likeness; and let him have dominion over triangles and over circles, and over spheres and over cubes, and over all of space, and every constructable subset that is constructed out of space.
27. So Euclid created a Mathematician in his own image, in the image of Euclid created he him; great and small created he them.
28. And Euclid blessed them, and Euclid said unto them, be fruitful, and multiply and prove theorems and subdue them: and have dominion over triangles, and over spheres, and over every set of points that cometh out of space.
29. And Euclid said Behold, I have given you every set bearing proof, which is on the face of all of space, and every set which is the fruit of a set yielding proof; to you it shall be for theorems.
31. And Euclid saw everything that he had made, and, behold, it was very good. And the evening and the morning were the sixth day.

Chapter 2

Thus space and all its subsets were finished, and all the uncountable host of them.
2. And on the seventh day Euclid ended the theory which he had made; and he rested on the seventh day from all the sets which he had made.
4. These are the generations of the sets and points and of the space when they were created, in the day that the Greek Euclid made the space and its subsets.
8. And Euclid builded a library eastward in Erewhon; and there he put the Mathematician whom he had formed.
9. And out of the points made the Greek Euclid to grow every set that is pleasant to the sight, and good for proof; the set of logic also in the midst of the library, and the set of knowledge of the excluded middle.

15. And Euclid took the Mathematician, and put him in the library to read it and keep it.

16. And Euclid commanded the Mathematician, saying, Of every set in the library thou may freely partake:

17. But of the set of the knowledge of the excluded middle, thou shalt not partake of it: for the day that thou partake thereof thou shalt surely die.

18. And Euclid said it is not good that the Mathematician should be alone; I will make a help meet for him.

21. And Euclid caused a deep sleep to fall on the Mathematician, and he slept; and he took *one* of his brains, and closed up the flesh instead thereof;

22. And the brain which the Greek Euclid had taken from Mathematician, made he a Subman and brought him unto the Mathematician.

23. And the Mathematician said, this is now bone of my bones, and flesh of my flesh; he shall be called Subman because he was taken out of man.

24. Therefore shall a man leave his mentors, and shall cleave to his student: and they shall be one mind.

25. And they were both stupid, the man and his student, and were not ashamed.

Chapter 3

Now the Uncountable was more subtle than any set of the space which the Greek Euclid had made. And he said unto the Subman, Yea, hath Euclid said, Ye shall not partake of every set in the library?

2. And the student said to the Uncountable, We may partake of all sets in the library:

3. But of the fruit of the set which is in the midst of the library, Euclid hath said, Ye shall not partake of it, neither shall ye touch it, lest ye die.

4. And the Uncountable said unto the student, Ye shall not surely die:

5. For Euclid doth know that in the day thou partake thereof, then your eyes shall be opened, and ye shall be as gods, knowing consistency and contradiction.

6. And when the student saw that the set *was* good for thought, and that it *was* pleasant to the mind, and a set to be desired to make one wise, he partook of the set thereof, and did think, and gave also unto his master with him; and he did partake.

7. And the eyes of both of them were opened, and they knew that they were stupid; and they invented classes and made systems.

8. And they heard the voice of Euclid walking in the library; and the Mathematician and his student hid themselves from the presence of Euclid amongst the footnotes in the library.

9. And Euclid called to the Mathematician and said unto him, Where art thou?

10. And he said, I heard thy voice in the library and I was afraid because I was stupid; and I hid myself.

11. And he said, Who told thee that thou wast stupid? Hast thou partaken of the set, whereof I forbade thee?

12. And the Mathematician said, The Subman whom thou gavest to be with me, he gave me of the set, and I did think.

13. And the Greek Euclid said unto the Subman, What is this that thou hast done? And the Subman said, The Uncountable beguiled me and I did think.

14. And Euclid said unto the Uncountable, Because thou hast done this thou art cursed above all other sets. To the foundations must thou go for all the days of thy life.

15. And I will put enmity between thee and the student, and between thy subsets and his subsets; it shall bruise thy pride, and thou shalt bruise his mind.

16. Unto the student he said, I will greatly multiply thy sorrow; in sorrow shalt thou bring forth lemmas; and thy desire shall be to thy master, and he shall rule over thee.

17. And unto the Mathematician he said, Because thou hast hearkened to the voice of thy student, and has thought of the set, of which I commanded thee, saying, Thou shalt not think of it: cursed is the structure for thy sake; in sorrow shalt thou think of it all the days of thy life.

22. And the Greek Euclid said, Behold, the Mathematician is become as one of us, to know consistency and contradiction: and now lest he put forth his hand and partake of the set of logic, and think, and prove theorems forever:

23. Therefore Euclid sent him forth from the library of Erewhon, to prove, without a library card, the theorems from whence he was taken.

24. So he drove out the Mathematician; and he placed at the east of the library of Erewhon Cher-u-bims Kronecker, and Hilbert, and Brouwer, with flaming words which turn every way, to keep the way of the set of logic.

PROBLEMS AND SOLUTIONS

EDITED BY EMORY P. STARKE

All problems (both elementary and advanced) proposed for inclusion in this Department should be sent to E. P. Starke, 1000 Kensington Ave., Plainfield, NJ 07060. Proposers of problems are urged to enclose any solutions or information that will assist the editors. Ordinarily, problems in well-known textbooks and results in generally accessible sources are not appropriate for this Department. No solutions (except those accompanying proposals) should be sent to Professor Starke.

ELEMENTARY PROBLEMS

Solutions of Elementary Problems should be sent to Problems Group, Mathematics Department, University of Maine, Orono, ME 04473. To facilitate their consideration, solutions of Elementary Problems in this issue should be typed (with double spacing) and should be mailed before May 31, 1970. Contributors (in the United States) who desire acknowledgement of receipt of their solutions are asked to enclose self-addressed stamped postcards.

An asterisk (*) means neither the proposer nor the editors supplied a solution.

E 2215. *Proposed by Michael Goldberg, Washington, D. C.*

Find the shapes of the quadrilaterals of least area K which enclose three given circles of radii 1, b, b.

E 2216. *Proposed by M. S. Klamkin, Ford Scientific Laboratory*

Which of the two integrals

$$\int_0^1 x^x dx, \quad \int_0^1 \int_0^1 (xy)^{xy} dx dy$$

is larger?

E 2217.* *Proposed by Robert Spira, Michigan State University*

A, B, C, D are complex numbers with $AC \neq 0$. Show that

$$\frac{\max(|AC|, |AD+BC|, |BD|)}{\max(|A|, |B|)\cdot\max(|C|, |D|)} \geq \frac{-1+\sqrt{5}}{2}.$$

E2218. *Proposed by CHED, Stratford-on-Penobscot*

An interesting geometric example of a noncommutative, nonassociative hoop is given by the binary operation $*$ that maps points A and B in the plane into point C in the plane where ABC is a counterclockwise-oriented equilateral

triangle. Prove the medial property for a hoop, that $(A*B)*(C*D)=(A*C)*(B*D)$ for all points A, B, C, D in the plane.

E 2219. *Proposed by Dan Marcus, York University, Toronto*

Let X be an unbounded set of real numbers. Show that

$$\{t \mid tX \text{ is dense modulo } 1\}$$

is dense in the reals.

E 2220. *Proposed by Otto Morphy*

Attach an integer $\theta(G)$ to each finite group G by means of the equations

$$\theta(I) = 1, \qquad \sum_{H\in G} \theta(H) = \operatorname{ord}(G),$$

where I is the trivial group and the sum ranges over all subgroups H of G, including I and G itself. This may be regarded as a recursion formula for θ so that there is precisely one such function. Show that $\theta(G)=0$ if G is nonabelian.

SOLUTIONS OF ELEMENTARY PROBLEMS

Unions of Subsets of a Finite Set

E 2106 [1968, 779; 1969, 697]. *Proposed by Bernt Lindström, University of Stockholm, Sweden.*

Let S be a set with n elements and $M_1, M_2, \cdots, M_{n+1}$ be nonempty subsets of S. Prove that one can find r, s, and $r+s$ distinct indices $i_1, i_2, \cdots, i_r, j_1, j_2, \cdots, j_s$ such that

$$M_{i_1} \cup \cdots \cup M_{i_r} = M_{j_1} \cup \cdots \cup M_{j_s}.$$

NOTE. The published solution is incorrect, as noted by E. F. Schmeichel. E. J. Zirkel, and others: if $M\cup M_1 = M\cup M_2$, it does not follow that $M_1=M_2$,

II. *Solution by Eddy Smet, University of Western Ontario.* Let $S=\{a_1, a_2, \cdots, a_n\}$ and let $M_1, \cdots, M_{n+1}$ be nonempty subsets of S. Corresponding to each set M_r is a nonzero vector

$$m_r = (m_{r1}, m_{r2}, \cdots, m_{rn}), \quad \text{where} \quad m_{ri} = \begin{cases} 1 & \text{if } a_i \in M_r \\ 0 & \text{if } a_i \notin M_r. \end{cases}$$

The $n+1$ vectors $m_1, \cdots, m_{n+1}$ are linearly dependent, and so, since $m_{ri}\geqq 0$, there are $r+s$ distinct indices $i_1, i_2, \cdots, i_r, j_1, j_2, \cdots, j_s$ ($r\geqq 1$, $s\geqq 1$, $n+1 \geqq r+s$), and positive numbers $\lambda_{i_1}, \cdots, \lambda_{i_r}, \lambda_{j_1}, \cdots, \lambda_{j_s}$ such that

$$\lambda_{i_1}m_{i_1} + \cdots + \lambda_{i_r}m_{i_r} = \lambda_{j_1}m_{j_1} + \cdots + \lambda_{j_s}m_{j_s},$$

which implies

$$M_{i_1} \cup \cdots \cup M_{i_r} = M_{j_1} \cup \cdots \cup M_{j_s}.$$

A Convergent Sequence

E 2115 [1968, 897; 1969, 828]. *Proposed by K. M. Brown, Cornell University*

Let a sequence S_n be defined by

$$S_n = \frac{n+1}{2^{n+1}} \sum_{i=1}^{n} 2^i/i, \qquad n = 1, 2, \cdots.$$

Show that $\lim_{n\to\infty} S_n$ exists and find the value of this limit.

Editorial Note. It has been brought to our attention that the proposed formula is an immediate consequence of

$$\lim_{n\to\infty} \frac{n}{x^n} \sum_{i=1}^{n} \frac{x^{i-1}}{i} = \frac{1}{x-1}, \qquad x > 1,$$

given on p. 450 of Randolph, *Basic Real and Abstract Analysis.*

Integers All of Whose Divisors are of the Same Form

E 2167 [1960, 413]. *Proposed by Steven Feigelstock, Polytechnic Institute of Brooklyn.*

Which positive integers of the form p^n-1, p a prime, have all their divisors of the same form?

Solution by D. C. B. Marsh, Colorado School of Mines. Let p^n-1 (p a prime) be an integer all of whose divisors have the same form.

1) If 2^a divides p^n-1, then $a\leqq 4$ since 2^5+1 is not a prime power.

2) If q^b divides p^n-1 (q an odd prime), then $q=2^k-1$, a Mersenne prime; furthermore, $b=1$ since q^2+1 is even with at least one nontrivial odd factor (and thus not a prime power).

3) Should q and r be two Mersenne primes, not both can divide p^n-1 since $qr+1$ is even with at least one nontrivial odd factor (and thus not a prime power).

4) If 2 and q (a Mersenne prime) both divide p^n-1, then $2q+1$ must also be a Mersenne prime, and thus $q=3$.

Thus, numbers of the desired form are limited to Mersenne primes and the divisors of 48 $(=2^4\cdot 3)$, each of which does possess the desired property.

Also solved by Orin Chein, W. F. Fox, D. E. Frohardt, J. R. Kuttler, C. B. A. Peck, Bob Prielipp, E. F. Schmeichel, Philip Trauber, C. S. Venkataraman (India), and the proposer.

A Correct Conjecture

E 2172 [1969, 553]. *Proposed by S. I. Drobnies, San Diego College*

Prove or disprove: The complex number z belongs to the set $\{\omega: |\omega| - \text{Re}\,\omega \leqq \frac{1}{2}\}$ if and only if z is a product ac such that $|\bar{c}-a| \leqq 1$.

Solution by J. R. Kuttler, The Johns Hopkins University Applied Physics Laboratory. The statement is correct and is a consequence of the identity

$$|ac| - \operatorname{Re} ac = \tfrac{1}{2}|\bar{c} - a|^2 - \tfrac{1}{2}(|c| - |a|)^2.$$

The "if" part is immediate. For the "only if" part choose a and c such that $|a| = |c| = |z|^{1/2}$ and the second term on the right vanishes.

Also solved by L. W. Beineke, Stephen Berman, J. C. Binz (Switzerland), M. T. Bird, Douglas Campbell, W. O. Egerland, E. N. Fischman, Michael Goldberg, D. L. Grant, Robert Heller, C. V. Heuer & G. A. Heuer, E. F. Knapp, D. C. B. Marsh, K. A. Ribet, Steve Rohde, Perry Scheinok, E. F. Schmeichel, J. H. Simester, Charles Snygg, John Swetits, Charles Wexler, Mark Yu, D. A. Zave, E. J. Zirkel, and P. J. Zwier.

The Ballot Problem Revisited

E 2174 [1969, 553]. *Proposed by Klaus Steffen, Johannes Gutenberg University, Mainz, Germany*

Consider the set M of arrays in a row of the $2n$ symbols $a_1, \cdots, a_n, b_1, \cdots, b_n$ such that a_i precedes a_{i+1} and b_i precedes b_{i+1} for $1 \leqq i \leqq n-1$. An inversion is a pair (a_i, b_i) such that a_i does not precede b_i. For $0 \leqq k \leqq n$, let v_k be the number of elements of M with k inversions. Prove or disprove that $v_0 = v_1 = \cdots = v_n$.

Solution by Harry Lass, Jet Propulsion Laboratory, California Institute of Technology. The set M consists of $\binom{2n}{n}$ arrays and is equivalent to the set of paths of the $n \times n$ grid problem wherein one can only move horizontally and vertically to reach an opposite vertex (the successive horizontal paths are denoted by $a_1, \cdots, a_n$, and the successive vertical paths are denoted by $b_1, \cdots, b_n$).

Clearly, the number of inversions is simply the number of vertical paths above the principal diagonal. Thus

$$v_0 = v_n = \binom{2n}{n} \Big/ (n+1),$$

obtained from the solution of the ballot problem, allowing ties. (See, e.g., Feller, *An Introduction to Probability Theory and its Applications*, vol. I.)

Now let $v_k(n)$ be the number of paths with k vertical paths above the principal diagonal (k inversions). If the first move is horizontal (a_1 recorded first), the diagonal will be reached after $2r$ moves in

$$\binom{2r-1}{r} \Big/ (2r-1) \text{ ways,}$$

and k inversions will be needed in the remaining $(2n-2r)$ moves accomplished in $v_k(n-r)$ ways. If the first move is vertical, the diagonal will be reached after $2s$ moves yielding s inversions, and $k-s$ further inversions are required. Thus

$$v_k(n) = \sum_{r=1}^{n-k} \frac{1}{2r-1}\binom{2r-1}{r} v_k(n-r) + \sum_{s=1}^{k} \frac{1}{2s-1}\binom{2s-1}{s} v_{k-s}(n-s).$$

By direct enumeration it is seen that $v_k(1)$, $v_k(2)$ are independent of k. Assume that $v_k(m)$ is independent of k for $m=1, 2, \cdots, n-1$, with

$$v_k(m) = \binom{2m}{m} \Big/ (m+1).$$

From above we obtain

$$v_{k+1}(n) - v_k(n) = \frac{1}{2k+1}\binom{2k+1}{k+1}\frac{1}{n-k}\binom{2n-2k-2}{n-k-1} - \frac{1}{2n-2k-1}\binom{2n-2k-1}{n-k}\frac{1}{k+1}\binom{2k}{k} \equiv 0.$$

Thus, by mathematical induction, we have shown that

$$v_0(n) = v_1(n) = v_2(n) = \cdots = v_n(n) = \binom{2n}{n} \Big/ (n+1).$$

Also solved by Michael Goldberg, M. G. Greening (Australia), and E. F. Schmeichel.

A Maximization Problem

E 2175 [1969, 554]. *Proposed by Harry Pollard, Purdue University*

Find the maximum value of the function

$$\exp\left(-\frac{1}{x}e^{-x}\right) + \exp(-xe^{-1/x}), \qquad 0 < x < \infty.$$

Solution by Simeon Reich, The Technion, Israel Institute of Technology. The maximum of our function $f(x)$ is attained at $x=1$, and is equal to $2e^{-1/e} = 1.38440$ approximately.

Proof. Since $f(x) = f(1/x)$, we may consider only the interval $[1, \infty)$. Clearly it is sufficient to prove that $f'(x) < 0$ for $x > 1$.

After some computation it is seen that we have to show that for $x>1$, $g(x)>0$, where $g(x) = \log x + x + (1/x)e^{-x} - 1/x - xe^{-1/x}$. We have $g(1)=0$. We shall prove that $g'(x)>0$ for $x>1$.

In order to show this we have to establish that

$$(1/x^2)e^{-x} + (1/x)e^{-x} + e^{-1/x} + (1/x)e^{-1/x} < 1/x + 1 + 1/x^2.$$

This is true because for $1 < x < \infty$, $e^{-1/x} < 1$, and $(1/x)e^{-1/x} < 1/x$. Also $(1/x^2)e^{-x} + (1/x)e^{-x} < 1/x^2$ because this is equivalent to $e^x > 1+x$ which holds for all $x \neq 0$.

Also solved by M. J. Brown, W. O. Egerland, Gretchen Ehman, M. A. Ettrick, R. D. Gee, A. F. Gentzel, Jr., Michael Goldberg, Emil Grosswald, Carl Hammer, Stephen Hoffman, Dennis Kump, P. A. Lindstrom, Beatriz Margolis (Argentina), Gus Mavrigian, J. V. Michalowicz, M. H. Moore, C. B. A. Peck, Marlow Sholander, E. F. Schmeichel, J. R. Ventura, Jr., L. E. Ward, Sr., M. R. Wise, and P. H. Young.

ADVANCED PROBLEMS

All solutions of Advanced Problems should be sent to J. Barlaz, Rutgers—The State University, New Brunswick, N. J. 08903. Solutions of Advanced Problems in this issue should be typed (with double spacing) on separate, signed sheets and should be mailed before May 31, 1970. Contributors (in the United States) who desire acknowledgement of receipt of their solutions are asked to enclose self-addressed, stamped postcards. An asterisk (*) means neither the proposer nor the editors supplied a solution.

5714. *Proposed by Marlow Sholander, Case Western Reserve University*

Let O be the origin, let $ABCD$ be a line segment with endpoints on the coordinate axes, and let OBC be equilateral. Let $r=AB$, $s=BC$, and $t=CD$ be positive integers. A triple r, s, t has a dual t, s, r and a sum $r+s+t$. It is called primitive if $(r, s, t)=1$. Two triples are called dependent if one is proportional to the other or to its dual. Pairs of independent primitive triples are called pips. Two such triples may have an element in common or they may have a common sum.

(i) What primes are found as elements shared by pips?

(ii) What is the minimum sum shared by pips?

5715.* *Proposed by Anon, Erewhon-upon-Wabash*

Let $R=Z[a_1, a_2, a_3, b_1, b_2, b_3]$ with the defining relations $b_1b_2=b_2b_3=b_3b_1=0$, $a_1b_1+a_2b_2+a_3b_3=1$. Let M be the R-module generated by x_1, x_2, x_3 with relations $b_1x_1=b_2x_2=b_3x_3=0$. Show that M is free of rank two. Generalize.

5716. *Proposed by Harry Kesten, Cornell University*

Let $X_1, X_2, \cdots$ be a sequence of independent identically distributed random variables and $S_n=\sum_{i=1}^{n} X_i$. Show that

$$\limsup_{n\to\infty} \frac{|X_n|}{|S_{n-1}|} = \infty \quad \text{with probability 1,} \tag{1}$$

whenever $E|X_i|$, the expectation of $|X_i|$, is infinite. Use (1) to derive the following theorem of Chow and Robbins (Proc. Nat. Acad. Sci., 47 (1961) 330–335): When $E|X_i| = \infty$ then for any sequence $\{b_n\}$ of positive numbers either

$$\limsup_{n\to\infty} \frac{|S_n|}{b_n} = 0 \quad \text{with probability 1}$$

or

$$\limsup_{n\to\infty} \frac{|S_n|}{b_n} = \infty \quad \text{with probability 1.}$$

5717. *Proposed by W. H. Ruckle, Lehigh University*

Without using the axiom of choice (i.e., no Hamel basis) construct a continuum $\{X_r\colon 0<r\leqq 1\}$ of linear subspaces of Hilbert space H which has the

following properties: (a) X_r is dense in H for each r; (b) if $r<s$, $X_r \subset X_s$ and X_s/X_r has uncountable dimension; (c) $\bigcup_r X_r \neq H$.

5718.* *Proposed by M. T. Boswell and G. P. Patil, Pennsylvania State University*

Find all functions $g(x)$ which are complex-valued functions of a real variable satisfying

$$\frac{dg(cx)}{dx} = c(g(x+1) - g(x))$$

for $x>0$ and fixed $c>0$. The linear functions satisfy this equation.

5719. *Proposed by Roger Lyndon, University of Michigan*

For $k \geqq 0$, let S be the set of all numbers of the form

$$s = \sqrt{k \pm \sqrt{k \pm \cdots \pm \sqrt{k}}}$$

with arbitrary finite sequence of signs. If $k \geqq 2$, then all s in S are real. Prove: (1) if $k=2$, then S is dense in the interval $(0, +2)$; (2) if $k>2$, then S is dense in no real interval.

SOLUTIONS OF ADVANCED PROBLEMS

Correction. As one of the solvers of Problem 5614 [1969, 708], read Sidney Glusman, intsead of Sidney Glicksman.

Continuous Functions with Zero

5657 [1969, 200]. *Proposed by Erwin Just, Bronx (N. Y.) Community College*

Prove that there exists a real valued function f, which is defined on $(0, \infty)$, is continuous at an infinite number of points and has the property that for each x, $f(x)=0$ if and only if $f(2x)\neq 0$.

I. *Solution by M. S. Demos, Villanova University.* Set $f(x)=0$ when $2^{2k}<x \leqq 2^{2k+1}$ and $f(x)=1$ when $2^{2k-1}<x\leqq 2^{2k}$, where k is any integer (positive or negative). This f satisfies the conditions of the problem.

II. *Comment by Frank Meyer, Hamline University.* A related question is whether there exists a continuous function f defined on $(0, \infty)$ such that $f(x)=0$ if and only if $f(2x)\neq 0$. Actually f cannot be continuous everywhere. For suppose f is a function such that $f(x)=0$ if and only if $f(2x)\neq 0$. Then in any interval $(a, b)\subset(0, \infty)$ which contains points α, β, $f(\alpha)=0$ and $f(\beta)\neq 0$, there is a point γ (not necessarily distinct from α) and a sequence $\{x_n\}$ such that $f(\gamma)=0$, $f(x_n)\neq 0$ for all n, and $\lim_{n\to\infty} x_n=\gamma$. But then by hypothesis, $f(2\gamma)\neq 0$, $f(2x_n)=0$ for all positive integers n and so

$$\lim_{n\to\infty} f(2x_n) \neq f(2\gamma).$$

Hence f is discontinuous at 2γ, a contradiction.

Also solved by forty-eight other readers.
The question answered in Solution II above was also raised and resolved by the proposer.

Duals of Nonmetrizable Topological Spaces

5658 [1969, 309]. *Proposed by Stephen Gelbart, Princeton University*

Let E be a locally convex topological vector space, and E' its topological dual equipped with the weak topology $\sigma(E', E)$. Is it true that E nonmetrizable implies E' nonmetrizable?

Solution by Donald Hartig, University of California at Santa Barbara. Not necessarily. $E'[\sigma(E', E)]$ is metrizable if and only if there exists a countable subset F of E having the property that for each finite subset S of E there exists a finite subset S' of F such that the balanced, convex hull of S is contained in the balanced, convex hull of S'. This condition is clearly satisfied for any locally convex space of countable dimension. Hence, for example, if E denotes the locally convex direct sum of a countable number of copies of the reals, E is not metrizable but $E'[\sigma(E', E)]$ is metrizable.

Also solved by Joel Anderson, Stephen Berman, Joe Cross, P. J. Holewijn & P. van der Steen (Netherlands), D. R. Horner, D. R. Kerr, Jr., H. Pétard, Barry Simon, D. R. Sherbert, Alan Shuchat, and the proposer.

Iterates and Fixed Points of a Continuous Function

5660 [1969, 309]. *Proposed by W. G. Dotson, Jr., North Carolina State University*

(1) Show that if T is a continuous function from the reals to the reals and there is a number x such that the sequence

$$(*) \qquad \left\{\frac{x + T(x) + T^2(x) + \cdots + T^{n-1}(x)}{n}\right\}$$

of Cesàro means of the Picard iterates is bounded, then T has a fixed point.

(2) Find a continuous function T from the reals to the reals which has a unique fixed point p and such that if $x \neq p$ then the sequence (*) is bounded but does not converge to p.

Solution by Douglas Lind, Stanford University. (1) Suppose T has no fixed points. Then by the intermediate value theorem, either $Tx > x$ for all x, or $Tx < x$ for all x. If $Tx > x$, then $\{T^n x\}$ is an increasing sequence. If $T^n x \to y$, then $T^{n+1}x \to Ty$, which implies $Ty = y$, a contradiction. Hence $T^n x \to \infty$, so the Cesàro means (*) are unbounded. The case $Tx < x$ is handled similarly.

(2) Let $Tx = -x+1$ $(x \leqq -1)$, $Tx = -2x(|x| \leqq 1)$, $Tx = -x-1$ $(x \geqq 1)$. Then if $x \geqq 1$,

$$\frac{1}{n}(x + Tx + \cdots + T^{n-1}x) = \begin{cases} -\frac{1}{2} & (n \text{ even}) \\ \frac{1}{2} + \frac{1}{n}(x - \frac{1}{2}) & (n \text{ odd}), \end{cases}$$

while if $x \leqq -1$,

$$\frac{1}{n}(x + Tx + \cdots + T^{n-1}x) = \begin{cases} \frac{1}{2} & (n \text{ even}) \\ -\frac{1}{2} + \frac{1}{n}(x + \frac{1}{2}) & (n \text{ odd}). \end{cases}$$

Since for $x \neq 0$ we have $|T^n x| \geqq 1$ for all sufficiently large n, we see that for $x \neq 0$ the sequence (*) is bounded but does not converge.

Also solved by David Boyd, Ted Cullen, R. V. Fuller, D. A. Hejhal, D. R. Horner, Emmett Keeler & Joel Spencer, B. G. Klein, O. P. Lossers (Netherlands), Dan Marcus, D. E. Myers, P. J. Owens, (England), Nicholas Passell, H. Pétard, Joel Spencer, Alberto Torchinsky, D. A. Zave, P. J. Zwier, and the proposer.

REVIEWS

EDITED BY KENNETH O. MAY

COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER

Printed materials for review should be sent to Kenneth O. May, American Mathematical Monthly, Department of Mathematics, University of Toronto, Toronto 5, Canada. Films and correspondence relating to films should be sent to Seymour Schuster, Carleton College, Northfield, Minnesota 55057.

All unsigned material is written by the editors. A boldface capital **C** *in the margin indicates that a review is based in part on classroom use. Professors willing to write such a review should first inform the editor in order to avoid duplication.*

Calculus. By Lipman Bers. Holt, Rinehart and Winston, New York, 1969. xv+932 pp. \$13.95. (Telegraphic Review, August 1969.)

Among the flood of calculus books which have recently appeared, this is surely one of the most noteworthy. The author has brilliantly succeeded in giving a fresh, appealing treatment of the standard topics in elementary calculus for science and mathematics students. The book manages to convey the beauty, as well as the power and utility, of calculus. The style invites students to read the text, not merely (say) the worked examples. Notations are simple and natural, explanations clear and to the point. The choices of illustrative scientific applications, as well as the attitude toward numerical calculation, are quite up to date. The homework exercises are well done; they range from very easy to rather challenging.

THE AMERICAN MATHEMATICAL MONTHLY

(FOUNDED IN 1894 BY BENJAMIN F. FINKEL)

THE OFFICIAL JOURNAL OF

THE MATHEMATICAL ASSOCIATION OF AMERICA, INC.

VOLUME 77 NUMBER 3

CONTENTS

(*continued on inside cover*)

MARCH 1970

NOTICE TO AUTHORS

Articles should be typewritten on good quality paper, triple spaced with wide margins. Submit the original and a duplicate; keep a complete copy for protection against loss. Please follow the format used in current issues of the MONTHLY. The manuscript must make the author's intent clear to the printer. A paper accepted for publication by the editor but not acceptable to the printer will be returned to the author for retyping. Please use the American Mathematical Society *Manual for Authors of Mathematical Papers* as your guide and please conform to correct English usage. A *Manual for Monthly Authors* is available on request.

Prospective MONTHLY authors are advised to consult the Statement of Policy in the January 1969 issue, p. 2. Specialized research is usually unsuitable. Backlog: Main Articles 11 months, Math. Notes 10 months, Research Problems 6 months, Classroom Notes 6 months, Math. Education 6 months.

EDITORIAL CORRESPONDENCE: to HARLEY FLANDERS, American Mathematical Monthly, Purdue University, Lafayette, IN 47907; ADVERTISING CORRESPONDENCE: to RAOUL HAILPERN, Mathematical Association of America, SUNY at Buffalo, N. Y. 14214; CHANGE OF ADDRESS and SUBSCRIPTIONS: to A. B. WILLCOX, Mathematical Association of America, 1225 Connecticut Ave., N.W., Washington, D.C. 20036.

Annual dues for members of the Association (including a subscription to the American Mathematical Monthly) are $10.00. For nonmembers the subscription price is $15.00.

PUBLISHED BY THE ASSOCIATION at Washington, D. C., and Menasha, Wisconsin during the months of January, February, March, April, May, June–July, August–September, October, November, December.

Second-class postage paid at Washington, D. C., and additional mailing offices. Acceptance for mailing at special rate of postage provided for in the Act of February 28, 1925, embodied in Paragraph 4, Section 538, P.L. and R., authorized April 1, 1926.

PRINTED IN THE UNITED STATES OF AMERICA

AN INTRODUCTION TO DISTRIBUTIONS

JOHN HORVÁTH, University of Maryland

1. In the late 1920's the British physicist P. A. M. Dirac introduced the so-called delta-function which has the following properties [6, section 15]: δ is defined and continuous on the whole real line $\boldsymbol{R}$, $\delta(x)=0$ for $x\neq 0$, $\int_{-\infty}^{\infty}\delta(x)dx=1$, and if f is a continuous function defined on $\boldsymbol{R}$, the relation

$$f(a) = \int_{-\infty}^{\infty} f(x)\delta(a-x)dx \tag{1}$$

holds for $a\in\boldsymbol{R}$. Furthermore δ is not only continuous but infinitely differentiable, and if f is a k times continuously differentiable function defined on $\boldsymbol{R}$, then

$$f^{(k)}(a) = \int_{-\infty}^{\infty} f(x)\delta^{(k)}(a-x)dx \tag{2}$$

for any $a\in\boldsymbol{R}$. If we define the Heaviside function Y by $Y(x)=0$ if $x<0$ and $Y(x)=1$ if $x>0$, then

$$\delta(x) = Y'(x). \tag{3}$$

Finally Dirac lists the remarkable equation

$$\frac{d}{dx}\log x = \frac{1}{x} - i\pi\delta(x) \tag{4}$$

which plays a role in the quantum theory of collision processes.

2. It was quickly pointed out by mathematicians that from the point of view of "rigorous" mathematics all this is nonsense. Of course it was perfectly clear to Dirac himself that δ is not a function in the classical sense of the word, and what is important are not the values assumed by δ at the points x, but rather the way δ and the $\delta^{(k)}$ act as operators on the functions f, as in formulas (1) and (2). It took about 30 years to discover the mathematical foundations of a correct formulation of the definition and properties of the delta-function, and it turned out that Dirac's brilliant intuition was right in every instance.

3. It is relatively easy to give a precise meaning to equation (1). It suffices indeed to consider δ not as a point function but as a set function. We call a collection $\mathfrak{X}$ of subsets of $\boldsymbol{R}$, or more generally of the n-dimensional euclidean space $\boldsymbol{R}^n$, a *tribe* if it satisfies the following two conditions:

Professor Horváth received his Ph.D. in 1947 at the University of Budapest, working under L. Fejér and F. Riesz. He has held positions at the C.N.R.S. Paris, the Universidad de los Andes, Colombia, and the University of Maryland, and visiting positions at the University of Madrid and the Université de Nancy. His research in analysis includes the books *Aproximación y funciones casi-analíticas*, Madrid 1956, *Topological Vector Spaces and Distributions* I, Addison-Wesley, 1966, and *Introducción a la topología general*, Pan American Union 1969. *Editor.*

(a) If $(A_k)_{k\in N}$ is a sequence of sets belonging to $\mathfrak{X}$, then their union belongs to $\mathfrak{X}$.

(b) If $A\in\mathfrak{X}$, then its complement $\complement A$, i.e., the set of those points of R^n which do not belong to A, also belongs to $\mathfrak{X}$.

A real-valued function f defined on R^n is said to be *measurable* with respect to the tribe $\mathfrak{X}$, or simply $\mathfrak{X}$-measurable, if for any pair a, b of real numbers such that $a<b$ the set $\{x \mid a\leqq f(x)<b\}$ belongs to $\mathfrak{X}$.

A *positive measure* defined on the tribe $\mathfrak{X}$ is a function μ which associates with each set $A\in\mathfrak{X}$ a number $0\leqq\mu(A)\leqq\infty$, and is such that if (A_k) is a sequence of pairwise disjoint sets belonging to $\mathfrak{X}$, then

$$\mu\left(\bigcup_{k=0}^{\infty} A_k\right)=\sum_{k=0}^{\infty}\mu(A_k).$$

Let now f be a bounded, $\mathfrak{X}$-measurable function. For every finite increasing sequence α of real numbers: $a_1<a_2<\cdots<a_p$ such that $a_1\leqq f(x)<a_p$ write $A_k^\alpha=\{x \mid a_k\leqq f(x)<a_{k+1}\}$. Then the Lebesgue-Stieltjes *integral* $\int f(x)d\mu(x)$ of f with respect to the positive measure μ is defined as the least upper bound of the values $\sum_{k=1}^{p} a_k\cdot\mu(A_k^\alpha)$, where α runs through all possible sequences of the kind described above.

For any $a\in R^n$ we denote by τ_a the translation map $x\mapsto x+a$ from R^n onto itself. For any subset A of R^n the translate by a of A is the set $\tau_a(A)=A+a=\{x+a \mid x\in A\}$. The translate by a of a positive measure μ defined on the tribe $\mathfrak{X}$ is given by

$$\tau_a(\mu)(A)=\mu(\tau_{-a}A)$$

for every $A\subset R^n$ such that $\tau_{-a}A\in\mathfrak{X}$.

If A is any subset of R^n, we denote by $-A$ its reflection with respect to the origin, i.e., the set $-A=\{-x \mid x\in A\}$. The reflection $\check{\mu}$ of the positive measure μ is then defined by

$$\check{\mu}(A)=\mu(-A)$$

for every $A\subset R^n$ such that $-A\in\mathfrak{X}$.

The Dirac measure δ is defined on the tribe of all subsets of R^n by the following conditions:

$$\delta(A)=1 \quad \text{if} \quad 0\in A,$$
$$\delta(A)=0 \quad \text{if} \quad 0\notin A.$$

Clearly $\check{\delta}=\delta$, and for any $a\in R^n$ the measure $\tau_a(\delta)$ is given by

$$\tau_a(\delta)(A)=1 \quad \text{if} \quad a\in A,$$
$$\tau_a(\delta)(A)=0 \quad \text{if} \quad a\notin A.$$

If f is a continuous function defined on R^n, then the relation

$$f(a) = \int_{R^n} f(x)d(\tau_a(\check{\delta}))(x) \tag{5}$$

holds for $a \in R^n$. This is the rigorous form of Dirac's formula (1). Of course (5) looks much clumsier than (1), but with the concept of convolution we shall be able to rewrite (5) in a form whose elegance matches that of (1). It is to keep in line with the general definition of a convolution that we have written $\check{\delta}$ instead of δ in (5), where it is in fact superfluous.

4. In the early 1940s Nicolas Bourbaki and Henri Cartan singled out a certain class of measures and characterized them not as set functions but as operations which associate with certain functions ϕ the number $\mu(\phi) = \int \phi(x) d\mu(x)$.

If ϕ is any real- or complex-valued function defined on R^n, the support Supp ϕ of ϕ is the closure of the set $\{x \mid \phi(x) \neq 0\}$. We denote by $\mathcal{K}(R^n)$, or simply by $\mathcal{K}$, the vector space consisting of all real-valued continuous functions ϕ on R^n such that Supp ϕ is a compact subset of R^n. A (real) *Radon measure* on R^n is a map μ from $\mathcal{K}$ into R which satisfies the following conditions:

(a) μ is linear, i.e., for $\alpha, \beta \in R$ and $\phi, \psi \in \mathcal{K}$ we have

$$\mu(\alpha\phi + \beta\psi) = \alpha\mu(\phi) + \beta\mu(\psi).$$

(b) For every compact subset K of R^n there exists a number $M_K > 0$ such that

$$|\mu(\phi)| \leqq M_K \max |\phi(x)|$$

whenever Supp $\phi \subset K$.

A complex Radon measure is defined similarly by taking for $\mathcal{K}$ the vector space of complex-valued continuous functions with compact support and maps $\mu: \mathcal{K} \to C$ which are C-linear, i.e., satisfy condition (a) for $\alpha, \beta \in C$.

This suggested to Laurent Schwartz in 1944 the idea to replace $\mathcal{K}$ by the smaller space of all infinitely differentiable functions with compact support, and condition (b) by a less stringent one. He thus obtained a new class of objects which contains not only the Dirac measure δ but also its derivatives $\delta^{(k)}$, which are not measures any more.

5. It was in fact not formula (2) which led Schwartz to the discovery of distributions but a problem posed by G. Choquet and J. Deny [**3**]. If σ is a similitude of R^n and f any function defined on R^n, we define the function $\sigma^*(f)$ by $\sigma^*(f)(x) = f(\sigma^{-1}x)$. Let Ω be a nonempty open subset of R^n and f a continuous function defined on R^n. Denote by $\mathfrak{F}$ the set of restrictions to Ω of all functions $\sigma^*(f)$, where σ runs through the similitudes of R^n. For the case $n = 2$ Choquet and Deny proved that every continuous function defined on Ω can be approximated uniformly on every compact subset of Ω by linear combinations of elements of $\mathfrak{F}$ if and only if f is not polyharmonic, i.e., does not satisfy an equation $\Delta^p f = 0$, where Δ is the Laplace operator and p an integer $\geqq 1$. The necessity of the condition is obvious. Indeed, if f is polyharmonic, then so is every function in $\mathfrak{F}$ and also their linear combinations. Since the uniform limit of polyharmonic

functions is polyharmonic, not every continuous function on Ω can be approximated by linear combinations of elements of $\mathfrak{F}$.

In their proof Choquet and Deny used the expansion of f into a series of harmonic polynomials. This tool was not powerful enough to yield the result for $n>2$. In a four-page-long paper [**19**] Schwartz gave a very elegant proof of the general theorem for any $n \geqq 2$, and, more importantly, discovered the theory of distributions practically overnight.

Besides the new point of view to consider Radon measures, Schwartz was also helped in his discovery by his familiarity with topological vector spaces. Indeed, in 1942 Dieudonné published a very important paper on the duality of normed vector spaces [4], and just as an exercise Schwartz generalized its results to Frechet spaces, i.e., locally convex, metrizable, and complete vector spaces. He never intended to publish these generalizations but they came extremely handy when the need for them arose.

6. We denote by $\mathfrak{D}(\boldsymbol{R}^n)$, or simply by $\mathfrak{D}$, the vector space of all real- or complex-valued functions ϕ defined on $\boldsymbol{R}^n$ which have continuous partial derivatives of all orders and are such that Supp ϕ is a compact subset of $\boldsymbol{R}^n$. It is not obvious that such functions exist which are not identically zero, but they can be constructed easily from Cauchy's infinitely differentiable function which assumes the value $\exp(-1/t^2)$ for $t>0$, and the value zero for $t \leqq 0$.

If $\alpha = (\alpha_1, \cdots, \alpha_n)$ is any n-tuple of positive integers, we shall denote by $|\alpha|$ the integer $\alpha_1 + \cdots + \alpha_n$ and by ∂^α the partial differential operator

$$\frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \cdots \partial x_n^{\alpha_n}}$$

A real (or complex) *distribution* on $\boldsymbol{R}^n$ is a map T from $\mathfrak{D}$ into $\boldsymbol{R}$ (or into $\boldsymbol{C}$, respectively), which satisfies the following two conditions:

(a) T is linear.

(b) For every compact subset K of $\boldsymbol{R}^n$ there exists a number $M_K>0$ and an integer $m(K) \geqq 0$ such that

$$|T(\phi)| \leqq M_K \max_{|\alpha| \leqq m(K)} \max_{x \in R^n} |\partial^\alpha \phi(x)|,$$

for every $\phi \in \mathfrak{D}$ which verifies Supp $\phi \subset K$.

Instead of $T(\phi)$ we shall mostly write $\langle T, \phi \rangle$.

Examples. (i) Let f be a continuous (or just locally Lebesgue-integrable) function defined on $\boldsymbol{R}^n$. We associate with it a distribution T_f defined by

$$\langle T_f, \phi \rangle = \int_{R^n} f(x)\phi(x)dx.$$

Here we can choose $M_K = \int_K |f(x)| dx$ and $m(K) = 0$ for every compact subset K of $\boldsymbol{R}^n$.

(ii) The Dirac measure δ is the distribution defined by

$$\langle \delta, \phi \rangle = \phi(0)$$

for every $\phi \in \mathfrak{D}$.

(iii) More generally, according to our definitions, every Radon measure defines a distribution for which $m(K) = 0$ for any compact $K \subset \boldsymbol{R}^n$. The Radon measures can be considered as the mathematical formulation of the physical concepts of mass or charge distributions.

(iv) Let Σ be a smooth surface, $d\sigma$ the surface element of Σ, denote by d/dn the derivation in the direction of the normal of Σ, and let f be a continuous function defined on Σ. Then

$$\langle T, \phi \rangle = \int_{\Sigma} f(x) \frac{d\phi(x)}{dn} d\sigma(x)$$

defines a distribution T which can be considered as a magnetic dipole layer on Σ, where f is the density of the magnetic moment. In this example we have to take $m(K) = 1$.

7. For any compact subset K of $\boldsymbol{R}^n$ denote by $\mathfrak{D}(K)$ the vector space of all functions which have continuous partial derivatives of all orders, and whose support is contained in K. Clearly $\mathfrak{D}$ is the union of all the spaces $\mathfrak{D}(K)$.

For any n-tuple of positive integers α we have a semi-norm on $\mathfrak{D}(K)$, i.e., a map $\phi \mapsto p_\alpha(\phi)$ defined by

$$p_\alpha(\phi) = \max_{x \in \boldsymbol{R}^n} | \partial^\alpha \phi(x) | .$$

The family (p_α) of all these semi-norms defines a topology on $\mathfrak{D}(K)$: a fundamental system of neighborhoods of $\phi \in \mathfrak{D}(K)$ is given by all finite intersections of sets of the form $\{\psi | p_\alpha(\psi - \phi) \leqq \epsilon\}$, as α runs through all n-tuples and ϵ through all strictly positive numbers. This topology is compatible with the vector space structure of $\mathfrak{D}(K)$, i.e., the two algebraic operations on $\mathfrak{D}(K)$, addition and multiplication by a scalar, are continuous. In particular, the topology of $\mathfrak{D}(K)$ is completely characterized by the system of neighborhoods of the zero element of $\mathfrak{D}(K)$: the neighborhoods of any $\phi \in \mathfrak{D}(K)$ are all the sets $V + \phi = \{\psi + \phi | \psi \in V\}$, where V is an arbitrary neighborhood of $0 \in \mathfrak{D}(K)$.

For any compact subset K of $\boldsymbol{R}^n$ we have a canonical map j_K from $\mathfrak{D}(K)$ into $\mathfrak{D}$, which associates with $\phi \in \mathfrak{D}(K)$ the same ϕ considered as an element of $\mathfrak{D}$. One then makes a topological vector space out of $\mathfrak{D}$ by putting on it the finest locally convex topology for which all the maps j_K are continuous. As we have observed above, to describe this topology it is sufficient to give the neighborhoods of the zero element of $\mathfrak{D}$. We say that a subset V of $\mathfrak{D}$ is balanced if for every $\phi \in V$ and every scalar t such that $|t| \leqq 1$, the element $t\phi$ also belongs to V; we say that V is convex if for $\phi \in V$, $\psi \in V$ and $0 \leqq t \leqq 1$ we have $(1-t)\phi + t\psi \in V$. A fundamental system of neighborhoods of $0 \in \mathfrak{D}$ is given by all balanced, convex sets V such that for every compact subset K of $\boldsymbol{R}^n$ the set $j_K^{-1}(V) = V \cap \mathfrak{D}(K)$ is a neighborhood of 0 in $\mathfrak{D}(K)$.

Once this topology is put on the space $\mathfrak{D}$, the distributions are simply the continuous linear maps from $\mathfrak{D}$ into the field of scalars, i.e., the continuous linear forms (or functionals). The space of all distributions on $\boldsymbol{R}^n$ is therefore the topological dual $\mathfrak{D}'$ of $\mathfrak{D}$.

We can define various topologies and modes of convergence on $\mathfrak{D}'$. Fortunately, for sequences all these definitions coincide, and we can say that a sequence (T_k) of distributions converges to the distribution T if for every $\phi \in \mathfrak{D}$ the sequence $(\langle T_k, \phi\rangle)$ converges to $\langle T, \phi\rangle$.

Understandably, the discovery of distributions has aroused great interest in the theory of topological vector spaces. Before 1950 functional analysts studied mainly normed spaces, and only a handful of papers were written on spaces which are not normable. In 1950 the basic paper of Dieudonné and Schwartz [**5**] was published, whose avowed purpose was to prove those results which are necessary for the theory of distributions. This was followed the next year by an equally basic paper of Bourbaki [**1**], and a few years later Grothendieck published his very deep results [**9**, **10**]. Since then the number of articles concerned with topological vector spaces has gone into the thousands. An excellent account of these developments can be found in Köthe's book [**14**].

On the other hand, several approaches to the theory of distributions have been devised, which avoid the explicit use of topological vector spaces. These so-called elementary approaches are, however, not very powerful and in particular they do not yield the applications to the theory of partial differential equations, some of which we shall mention below, which are the strongest justifications of the theory of distributions.

8. The classical results concerning normed spaces, when generalized to topological vector spaces, can be used to prove results concerning distributions. We shall illustrate this by two examples.

(a) By example (i) above, $\mathfrak{D}$ can be considered as a linear subspace of the space $\mathfrak{D}'$ of all distributions, if we identify $\phi \in \mathfrak{D}$ with the distribution T_ϕ. Then $\mathfrak{D}$ is a dense subspace of $\mathfrak{D}'$. *Proof.* If $\mathfrak{D}$ were not dense in $\mathfrak{D}'$, then by the Hahn-Banach theorem there would exist a continuous linear form L on $\mathfrak{D}'$ which is not identically zero but $L(\phi) = \langle L, \phi\rangle = 0$ for every $\phi \in \mathfrak{D}$ [**13**, prop. 1.6.2, p. 49]. Now it is easy to see that the space $\mathfrak{D}$ is reflexive, i.e., for every continuous linear form L on $\mathfrak{D}'$ there exists $\psi \in \mathfrak{D}$ such that $\langle L, T\rangle = \langle T, \psi\rangle$ for every $T \in \mathfrak{D}'$ [**13**, example 3.9.6, p. 241]. If $\langle L, \phi\rangle = \langle T_\phi, \psi\rangle = 0$ for every $\phi \in \mathfrak{D}$, then in particular $\langle L, \bar{\psi}\rangle = \int |\psi(x)|^2 dx = 0$, hence $\psi = 0$ and therefore $L = 0$. Thus $\mathfrak{D}$ must be dense in $\mathfrak{D}'$.

(b) The classical Banach-Steinhaus theorem [**13**, prop. 1.8.1, p. 63] can be generalized to the space $\mathfrak{D}$ to yield the following result: Let (T_k) be a sequence of distributions such that for every $\phi \in \mathfrak{D}$ the sequence of scalars $(\langle T_k, \phi\rangle)$ converges to some number $T(\phi)$. Then the map $T: \phi \mapsto T(\phi)$ is a distribution, i.e., a continuous linear form on $\mathfrak{D}$. It is fairly obvious that T is linear, the emphasis is on the continuity of T. Let us also observe that for Banach spaces the proof

of the Banach-Steinhaus theorem uses Baire's theorem, according to which if a complete metric space is the union of a sequence of closed subsets (F_k), then at least one set F_k has an interior point. Now in $\mathfrak{D}$ Baire's theorem does not hold, but the Banach-Steinhaus theorem is still true because $\mathfrak{D}$ is a so-called barrelled space. It was Bourbaki [1] who characterized those spaces for which the Banach-Steinhaus theorem holds by the property of barrelledness, which is usually easy to establish.

Example (v). The distribution v.p.$(1/x)$ on $\boldsymbol{R}$ is defined by

$$\left\langle \text{v.p.}\frac{1}{x}, \phi \right\rangle = \lim_{\epsilon\to 0} \int_{|x|>\epsilon} \frac{\phi(x)}{x}\, dx \tag{6}$$

for $\phi \in \mathfrak{D}(\boldsymbol{R})$. That it is a distribution follows from the fact that the right-hand side of (6) is also the limit of the sequence $(\int_{|x|>1/k} x^{-1}\phi(x)dx)$. The letters v.p. are used because the right-hand side of (6) is the Cauchy principal value (in French: *valeur principale*) of the integral $\int_{-\infty}^{\infty} x^{-1}\phi(x)dx$. The symbol $1/x$ in Dirac's formula (4) is meant to designate the distribution v.p.$(1/x)$.

9. Let us now examine how the various operations, which are classically performed on functions, generalize to distributions.

The first of these operations is *differentiation*. To motivate the definition, let f be a continuous function defined on the real line, and which has a continuous first derivative ∂f. It is natural to require that the derivative ∂T_f of the distribution T_f associated with f shall be the distribution $T_{\partial f}$ associated with ∂f. If this is so, then for every $\phi \in \mathfrak{D}(\boldsymbol{R})$ we have

$$\begin{aligned}\langle \partial T_f, \phi\rangle &= \langle T_{\partial f}, \phi\rangle = \int_{-\infty}^{\infty} \partial f(x)\phi(x)dx \\ &= [f(x)\phi(x)]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} f(x)\partial\phi(x)dx = -\langle T_f, \partial\phi\rangle.\end{aligned}$$

Let us write briefly ∂_j for the partial differential operator $\partial/\partial x_j$ $(1 \leqq j \leqq n)$ on $\boldsymbol{R}^n$. The above consideration leads us to define the partial derivative $\partial_j T$ of the distribution T on $\boldsymbol{R}^n$, by requiring that the formula

$$\langle \partial_j T, \phi\rangle = -\langle T, \partial_j \phi\rangle$$

be verified for every $\phi \in \mathfrak{D}(\boldsymbol{R}^n)$. We have to prove, of course, that $\partial_j T$ is a distribution. But the topology of $\mathfrak{D}$ has been so cleverly devised that $\phi \mapsto \partial_j\phi$ is a continuous linear map from $\mathfrak{D}$ into itself, and therefore the linear form $\phi \mapsto \langle T, \partial_j \phi\rangle$ is continuous.

More generally, if α is any n-tuple of positive integers, we have

$$\langle \partial^\alpha T, \phi\rangle = (-1)^{|\alpha|}\langle T, \partial^\alpha \phi\rangle$$

for every $\phi \in \mathfrak{D}$.

Example (vi). Let us show that the derivative of the distribution associated

with the Heaviside function Y is the Dirac measure δ. This justifies Dirac's formula (3). For any $\phi \in \mathfrak{D}(R)$ we have indeed

$$\langle \partial T_Y, \phi \rangle = -\langle T_Y, \partial \phi \rangle = -\int_0^\infty \partial \phi(x) dx$$
$$= -[\phi(x)]_0^\infty = \phi(0) = \langle \delta, \phi \rangle.$$

10. The second operation we can perform on distributions is the *multiplication* by an infinitely differentiable function. Again, if T_f is the distribution associated with the continuous function f, and ψ is an infinitely differentiable function defined on R^n, it is natural to require that ψT_f shall be the distribution associated with ψf. If this is so, then for every $\phi \in \mathfrak{D}$ we have

$$\langle \psi T_f, \phi \rangle = \langle T_{\psi f}, \phi \rangle = \int_{R^n} \psi(x) f(x) \phi(x) dx$$
$$= \int_{R^n} f(x) \psi(x) \phi(x) dx = \langle T_f, \psi \phi \rangle.$$

Thus if ψ is an infinitely differentiable function and T any distribution on R^n, we define ψT by asking that

$$\langle \psi T, \phi \rangle = \langle T, \psi \phi \rangle$$

be verified for every $\phi \in \mathfrak{D}$. This definition is justified by the fact that $\psi\phi$ has compact support, and $\phi \mapsto \psi \phi$ is a continuous map from $\mathfrak{D}$ into itself.

11. Perhaps the most important of all operations is the *convolution* of two distributions. First let f and g be two continuous functions on R^n, one of which has compact support. Their convolution $h = f * g$ is classically defined by

$$h(x) = \int_{R^n} f(x-y) g(y) dy. \tag{7}$$

If f is any function defined on R^n, the function $\check{f}$ is defined by $x \mapsto f(-x)$, and for any $a \in R^n$ the function $\tau_a(f)$ is defined by $x \mapsto f(x-a)$. With these notations formula (7) can be rewritten as

$$h(x) = \int_{R^n} (\tau_x \check{f})(y) g(y) dy.$$

It is therefore justified to define the convolution $T * \phi$ of a distribution T and a function $\phi \in \mathfrak{D}$ as the function given by

$$(T * \phi)(x) = \langle T, \tau_x \check{\phi} \rangle$$

for $x \in R^n$. One then proves that $T * \phi$ is an infinitely differentiable function [**13**, prop. 4.10.1, p. 402; **22**, Theorem 27.2, p. 287].

12. It will not be possible to define the convolution of two arbitrary distribu-

tions, hence we have to introduce a certain subclass of $\mathfrak{D}'$. We shall say that the distribution S has compact support if there exists a compact subset K of $\boldsymbol{R}^n$ such that $\langle S, \phi\rangle = 0$ whenever $\phi \in \mathfrak{D}$ is such that $K \cap \operatorname{Supp} \phi = \varnothing$.

If S is a distribution with compact support, then we can give a meaning to the symbol $\langle S, \psi\rangle$ for any infinitely differentiable function ψ. Indeed, let χ be a function belonging to $\mathfrak{D}$, which is equal to 1 on the compact set K which figured above, when we explained what it meant for S to have a compact support. Then $\chi\psi$ belongs to $\mathfrak{D}$ and the value of $\langle S, \chi\psi\rangle$ is independent of the choice of χ, since if χ_1 is a second function which has the same properties as χ, then the support of $(\chi-\chi_1)\psi$ does not meet K. Therefore we can set, by definition,

$$\langle S, \psi\rangle = \langle S, \chi\psi\rangle.$$

13. If T is an arbitrary distribution, we define $\check{T}$ by requiring that the relation

$$\langle \check{T}, \phi\rangle = \langle T, \check{\phi}\rangle$$

be satisfied for $\phi \in \mathfrak{D}$. This definition is justified by the fact that, if $T = T_f$ and $\check{T} = T_{\check{f}}$, then

$$\langle \check{T}, \phi\rangle = \int_{R^n} f(-x)\phi(x)dx = \int_{R^n} f(x)\phi(-x)dx = \langle T, \check{\phi}\rangle.$$

Let now S be a distribution with compact support, and T an arbitrary distribution. Their convolution $S * T$ is the distribution which verifies the relation

$$\langle S * T, \phi\rangle = \langle S, \check{T} * \phi\rangle$$

for every $\phi \in \mathfrak{D}$. It has to be proved, of course, that $S * T$ is indeed a distribution, i.e., that the map $\phi \mapsto \langle S, \check{T} * \phi\rangle$ is continuous [22, Theorem 27.3, p. 289].

To justify this definition, consider the case $S = T_f$, $T = T_g$, where f and g are continuous functions and f has compact support. Then we have

$$\begin{aligned}\langle S * T, \phi\rangle &= \langle S, \check{T} * \phi\rangle \\ &= \int_{R^n} f(x) \int_{R^n} g(-y)\phi(x-y)dy\,dx \\ &= \int_{R^n} f(x) \int_{R^n} g(z-x)\phi(z)dz\,dx \\ &= \int_{R^n} h(z)\phi(z)dz = \langle T_h, \phi\rangle,\end{aligned}$$

where $h = f * g$. Thus our definition yields $T_f * T_g = T_{f*g}$, which is reasonable to require.

14. We shall now list some properties of convolution.

(a) For any distribution T we have $\delta * T = T$. In other words, the Dirac measure is the unit element for the convolution considered as an algebraic opera-

tion. If T is the distribution associated with a continuous function f, then this is the promised elegant form of Dirac's relations (1) and (5).

(b) If S is a distribution with compact support, and T an arbitrary distribution, then

$$\partial_j(S * T) = (\partial_j S) * T = S * (\partial_j T)$$

for $1 \leqq j \leqq n$.

(c) Combining the preceding two relations, we obtain for any n-tuple of positive integers α, and for any distribution T, the relation

$$(\partial^\alpha \delta) * T = \partial^\alpha T,$$

which is the equivalent of Dirac's formula (2). Thus derivation has become a convolution with a fixed distribution.

(d) More generally, if $\xi = (\xi_1, \cdots, \xi_n)$ is a point of R^n and $\alpha = (\alpha_1, \cdots, \alpha_n)$ an n-tuple of positive integers, let us write ξ^α for $\xi_1^{\alpha_1}\xi_2^{\alpha_2} \cdots \xi_n^{\alpha_n}$. A polynomial of degree m in the n variables $\xi_1, \cdots, \xi_n$ can then be written as $P(\xi) = \sum_{|\alpha| \leq m} a_\alpha \xi^\alpha$, where the coefficients a_α are scalars. If for each j $(1 \leqq j \leqq n)$ we replace ξ_j by the differential operator ∂_j, we obtain the partial differential operator with constant coefficients $P(\partial) = \sum_{|\alpha| \leq m} a_\alpha \partial^\alpha$. For any such operator $P(\partial)$, and for any distribution T we have

$$P(\partial)\delta * T = P(\partial)T.$$

15. Let $P(\partial)$ be a partial differential operator with constant coefficients. Then

$$P(\partial)T = S, \tag{8}$$

where S is a given distribution and T an unknown distribution, is called a linear partial differential equation with constant coefficients. Though the theory of these equations belongs to the most classical parts of analysis, it has made very significant progress since the discovery of distributions. An account of these developments can be found in the books of Hörmander [12] and Trèves [23].

In the first place, distribution theory helped to clear up the concept of a fundamental (or elementary) solution of a partial differential operator with constant coefficients. It was always *felt* that, for instance, if $|x|$ denotes the length of the vector $x = (x_1, \cdots, x_n) \in R^n$ given by $|x|^2 = x_1^2 + \cdots + x_n^2$, then the function E defined by

$$E(x) = -\frac{\Gamma((n-2)/2)}{4\pi^{n/2}} \frac{1}{|x|^{n-2}}$$

is the fundamental solution of the Laplace operator

$$\Delta = \partial_1^2 + \partial_2^2 + \cdots + \partial_n^2$$

for $n \geqq 3$. The reasons for this belief were the following:

(a) E has a singularity at the origin, and outside $\{0\}$ it satisfies $\Delta E=0$;

(b) a solution u of the Poisson equation $\Delta u=f$ is given by the Newtonian potential

$$u(x) = (f * E)(x) = -\frac{\Gamma((n-2)/2)}{4\pi^{n/2}} \int_{R^n} \frac{f(y)}{|x-y|^{n-2}}\, dy.$$

With the help of distributions, a very simple and satisfactory definition of the fundamental solution of the operator $P(\partial)$ can be given: it is a distribution E which verifies $P(\partial)E=\delta$.

If E is a fundamental solution of $P(\partial)$, and if S is a distribution with compact support, then $T=S*E$ is a solution of (8).

Proof. $P(\partial)(S*E)=S*(P(\partial)E)=S*\delta=S$.

This theorem already shows the importance of the result, obtained in 1953 independently by Malgrange and Ehrenpreis, according to which every partial differential operator with constant coefficients has a fundamental solution.

16. We shall say that the distribution T is equal to the function f (or simply that T is f) in the open subset Ω of R^n, if $\langle T, \phi\rangle=\int_\Omega f(x)\phi(x)dx$ for every $\phi\in\mathfrak{D}$ such that Supp $\phi\subset\Omega$. We shall say that the partial differential operator $P(\partial)$ is hypoelliptic, if whenever $P(\partial)T$ is an infinitely differentiable function in the open set $\Omega\subset R^n$, then T itself is an infinitely differentiable function in Ω. It has been known for a long time that Δ, and, more generally, that every elliptic differential operator is hypoelliptic, and the problem arose to characterize the hypoelliptic operators.

It is fairly easy to prove that $P(\partial)$ is hypoelliptic if and only if it has a fundamental solution which is an infinitely differentiable function outside $\{0\}$. Actually, if $P(\partial)$ is hypoelliptic, then every fundamental solution is trivially infinitely differentiable outside $\{0\}$ since δ is equal to 0 and therefore infinitely differentiable outside $\{0\}$. But the interesting problem was to give a condition which uses only algebraic properties of the polynomial $P(\xi)$. This was done most brilliantly by Hörmander in his thesis in 1957.

If in $P(\xi)$ we replace the real n-tuple $\xi=(\xi_1, \cdots, \xi_n)$ by the complex n-tuple $\zeta=(\zeta_1, \cdots, \zeta_n)\in C^n$, where $\zeta_j=\xi_j+i\eta_j$, $\xi_j\in R$, $\eta_j\in R$ $(1\leqq j\leqq n)$, then we obtain the polynomial $P(\zeta)$ in n complex variables. We also set $\zeta=\xi+i\eta$, with $\xi=(\xi_1, \cdots, \xi_n)\in R^n$ and $\eta=(\eta_1, \cdots, \eta_n)\in R^n$. Hörmander's necessary and sufficient condition for the hypoellipticity of $P(\partial)$ is as follows: if $P(\zeta)=0$ and $|\zeta|\to\infty$, then necessarily $|\eta|\to\infty$.

17. The last operation we want to introduce is the *Fourier transformation* of distributions. If $x=(x_1, \cdots, x_n)$ and $\xi=(\xi_1, \cdots, \xi_n)$ are two vectors in R^n, we shall denote by $x\cdot\xi$ their inner product $x_1\xi_1+\cdots+x_n\xi_n$. The Fourier transform $\hat{f}=\mathfrak{F}f$ of an integrable function f on R^n is classically defined by

$$\hat{f}(\xi) = \int_{R^n} f(x)e^{-2\pi i x\cdot\xi}dx.$$

If $T = T_f$, and if we require that $\mathfrak{F}T = \hat{T} = T_{\hat{f}}$, then we have

$$\langle \mathfrak{F}T, \phi\rangle = \langle T, \mathfrak{F}\phi\rangle \tag{9}$$

for every $\phi \in \mathfrak{D}$, since both sides of (9) are equal to

$$\int_{R^n}\int_{R^n} f(x)\phi(\xi)e^{-2\pi i x\cdot\xi}dxd\xi.$$

One is therefore tempted to define the Fourier transform $\mathfrak{F}T$ of an arbitrary distribution T by (9). Unfortunately this cannot be done, since $\mathfrak{F}\phi$ is not in $\mathfrak{D}$, unless ϕ is identically zero. Indeed, by a theorem of Paley and Wiener, if ϕ has compact support, then $\mathfrak{F}\phi$ is the restriction to $\boldsymbol{R}^n$ of an entire function of exponential type on $\boldsymbol{C}^n$, and can therefore not have a compact support unless it vanishes identically.

There are two ways out of this dilemma.

18. The first of them has been used already by Schwartz himself in his book **[20]**. He introduces a vector space $\mathcal{S}$ which is larger than $\mathfrak{D}$, and which consists of all infinitely differentiable functions ϕ such that for every integer k and every n-tuple of positive integers α the expression

$$(1 + |x|^2)^k | \partial^\alpha\phi(x)| \tag{10}$$

tends to zero as $|x| \to \infty$. For each k and α we then define the seminorm $p_{k,\alpha}(\phi)$ as the maximum of (10) as x varies in $\boldsymbol{R}^n$, and the family of seminorms $(p_{k,\alpha})$ defines a topology on $\mathcal{S}$, which is compatible with the vector space structure of $\mathcal{S}$ (cf. **7**). It then follows from Fourier's inversion formula that $\phi \mapsto \hat{\phi}$ is an algebraical and topological isomorphism from $\mathcal{S}$ onto itself.

Next Schwartz considers the topological dual $\mathcal{S}'$ of $\mathcal{S}$, and calls its elements *tempered distributions*. Since the canonical injection from $\mathfrak{D}$ into $\mathcal{S}$ is continuous, and since $\mathfrak{D}$ is dense in $\mathcal{S}$, every tempered distribution is in particular a distribution, i.e., $\mathcal{S}'$ can be considered as a subset of $\mathfrak{D}'$. The Fourier transform $\mathfrak{F}T = \hat{T}$ of any $T \in \mathcal{S}'$ is then defined by requiring that (9) shall be satisfied for every $\phi \in \mathcal{S}$. Then $T \mapsto \hat{T}$ is a bijective map from $\mathcal{S}'$ onto itself, and for appropriate topologies on $\mathcal{S}'$ it is even bicontinuous.

Let us now list some properties of the Fourier transformation.

(a) The Fourier transform of the Dirac measure is the constant function equal to 1.

(b) For any n-tuple of positive integers α, the Fourier transform of $\partial^\alpha\delta$ is the monomial function $\xi \mapsto (2\pi i\xi)^\alpha$. For this reason it is more convenient to use the differential operators $D_j = (1/2\pi i)\partial_j$ rather than ∂_j $(1 \leq j \leq n)$. If we write a partial differential operator with constant coefficients in the form $P(D) = \sum_{|\alpha| \leq m} a_\alpha D^\alpha$, where $D^\alpha = D_1^{\alpha_1} \cdots D_n^{\alpha_n}$, then the Fourier transform of $P(D)\delta$ is the polynomial function $\xi \mapsto P(\xi) = \sum_{|\alpha| \leq m} a_\alpha \xi^\alpha$.

(c) If S is a distribution with compact support, and T a tempered distribution, then

$$\mathfrak{F}(S * T) = \mathfrak{F}(S) \cdot \mathfrak{F}(T). \tag{11}$$

Every distribution with compact support is tempered, and, by an extension of the Paley-Wiener theorem, its Fourier transform is the restriction to $\boldsymbol{R}^n$ of an entire function on $\boldsymbol{C}^n$. Therefore the right-hand side of (11) is well defined.

19. The second solution to our problem has been used mainly by Ehrenpreis and Gel'fand-Šilov. They introduce the vector space $\mathcal{Z}$ of all entire functions of exponential type on $\boldsymbol{C}^n$, whose restrictions to $\boldsymbol{R}^n$ belong to $\mathcal{S}$, and equip $\mathcal{Z}$ with an appropriate topology. Then $\phi \mapsto \hat{\phi}$ is an isomorphism from $\mathcal{Z}$ onto $\mathfrak{D}$ and for any $T \in \mathfrak{D}'$ we can define $\mathfrak{F}T$ by asking that (9) be satisfied for every $\phi \in \mathcal{Z}$. Then $\mathfrak{F}T$ will not be a distribution any more but a generalized distribution, i.e., an element of the topological dual of $\mathcal{Z}$. For an account of this theory we refer to the book of Gel'fand and Šilov [**8**], to the references given there, and to a paper by Ehrenpreis [**7**].

20. The considerations of no. **18** have led Schwartz to pose the problem of division of distributions: Given a polynomial $P(\xi)$ and a tempered distribution S, does there exist a tempered distribution X such that $P(\xi)X = S$?

If the answer to this problem is affirmative, then every linear partial differential equation $P(D)T = S$, where S is a given tempered distribution, has a tempered solution T. Indeed, there exists then a tempered distribution X such that $P(\xi)X = \hat{S}$, and X will be the Fourier transform of the required distribution T. In particular, every partial differential operator $P(D)$ has then a tempered fundamental solution.

After resisting several attempts, the problem of division was solved by Hörmander [**11**] and Łojasiewicz [**15**]. The latter even proved the possibility of division by a holomorphic function, and his methods were later used to study the structure of analytic manifolds. The solution of the problem of division inspired Malgrange to prove important theorems concerning overdetermined systems of partial differential equations [**16, 17**], and subsequently to obtain significant results in differential topology, in particular a proof of the so-called preparation theorem for differentiable functions, for which we refer to his book [**18**].

A lecture delivered at the session on the Theory of Distributions of the Mathematical Association of America held in New Orleans, January 27, 1969.

References

1. N. Bourbaki, Sur certains espaces vectoriels topologiques, Ann. Inst. Fourier (Grenoble), 2 (1950) 5–16.

2. ———, Eléments de mathématique, Livre V, Espaces vectoriels topologiques, Actualités Scientifiques et Industrielles, 1189, 1229, 1230, Hermann, Paris, 1966, 1955.

3. G. Choquet and J. Deny, Sur quelques propriétés de moyenne caractéristiques des fonctions harmoniques et polyharmoniques, Bull. Soc. Math. France, 72 (1944) 118–140.

4. J. Dieudonné, La dualité dans les espaces vectoriels topologiques, Ann. Sci. Ecole Norm. Sup., 59 (1942) 107–139.

5. J. Dieudonné and L. Schwartz, La dualité dans les espaces ($\mathfrak{F}$) et ($\mathfrak{LF}$), Ann. Inst. Fourier (Grenoble), 1 (1949) 61–101.

6. P. A. M. Dirac, The principles of quantum mechanics, 3rd ed., University Press, Oxford, 1947.

7. L. Ehrenpreis, Analytic functions and the Fourier transform of distributions, I, Ann. of Math., 63 (1956) 129–159; II, Trans. Amer. Math. Soc., 89 (1958) 450–483.

8. I. M. Gel'fand and G. E. Šilov, Generalized functions, II, Spaces of fundamental and generalized functions, Gosudarstvennoe izdatel'stvo fizikomatematičeskoĭ literatury, Moscow, 1958.

9. A. Grothendieck, Sur les espaces ($\mathfrak{F}$) et ($\mathfrak{DF}$), Summa Brasil. Math., 3 (1954) 57–123.

10. ———, Produits tensoriels topologiques et espaces nucléaires, Mem. Amer. Math. Soc., 16 (1955).

11. L. Hörmander, On the division of distributions by polynomials, Ark. Mat., 3 (1958) 555–568.

12. ———, Linear partial differential operators, Die Grundlehren der mathematischen Wissenschaften, vol. 116, Springer-Verlag, Berlin-Göttingen-Heidelberg, 1963.

13. J. Horváth, Topological vector spaces and distributions, I, Addison-Wesley, Reading, Mass., 1966.

14. G. Köthe, Topologische lineare Räume, I, Die Grundlehren der mathematischen Wissenschaften, vol. 107, Springer-Verlag, Berlin-Göttingen-Heidelberg, 1960.

15. S. Łojasiewicz, Sur le problème de division, Studia Math., 18 (1959) 87–136 or Rozprawy Mat., 22 (1961).

16. B. Malgrange, Sur les systèmes différentiels à coefficients constants, Les équations aux dérivées partielles (Paris, 1962), pp. 113–122. Editions du Centre National de la Recherche Scientifique, Paris, 1963.

17. ———, Systèmes différentiels à coefficients constants, Séminaire Bourbaki, 15e année: 1962/63, exposé 246, Secrétariat mathématique, Paris, 1963.

18. ———, Ideals of differentiable functions, Tata Institute of Fundamental Research, Study in Mathematics no. 3, University Press, Oxford, 1966.

19. L. Schwartz, Sur certaines familles non fondamentales de fonctions continues, Bull. Soc. Math. France, 72 (1944) 141–145.

20. ———, Théorie des distributions, 3rd ed., Publications de l'Institut Mathématique de l'Université de Strasbourg no. 9–10, Hermann, Paris, 1966.

21. ———, Some applications of the theory of distributions, Lectures on modern mathematics, I, pp. 23–58, Wiley, New York, 1963.

22. F. Trèves, Topological vector spaces, distributions and kernels, Academic Press, New York, 1967.

23. ———, Linear partial differential equations with constant coefficients, Gordon and Breach, New York, 1966.

A Remark Concerning the Alaoglu Theorem

ELLEN TORRANCE, Mount Holyoke College

Use a Tychonoff Theorem attack
To prove this Alaoglu fact:
  The closed unit ball
  In a normed space dual
Is Hausdorff, and weak-* compact.

APPLICATIONS OF DISTRIBUTIONS TO PDE THEORY

F. TREVES, Purdue University

The impact of Schwartz distributions on the theory of linear partial differential equations is today a well-recognized fact. There is no exaggeration in saying that the theory has been revolutionized during the fifties with the appearance of the works of Malgrange, Ehrenpreis, Hörmander, *et al.* Great advances have been made in the last fifteen years; further advances are still being made, with the theory of pseudodifferential operators, of hyperfunctions, of analytic functionals—in the wake of the first breakthroughs.

The advantages of distributions are manifold: a much needed increase in the *"reservoir"* of data and possible solutions (that the increase was needed is made clear by the "avant-garde" work of Hadamard, Sobolev, Bochner, and others); an automatization of the elementary operations such as taking limits, differentiations, convolutions, regularizations, Fourier transformation, etc. Another aspect, which I should like to emphasize here, has been the diffusion throughout PDE theory of the methods and results of functional analysis, particularly of the theory of general locally convex spaces. This theory was to a large extent built by S. Banach, but could not find significant applications until the late forties, when L. Schwartz gave his formulation of distribution theory. Not only did this formulation call for a systematic investigation of the properties of locally convex spaces (eventually presented in Bourbaki [1]) but it promised extensive applications. This, J. Dieudonné and L. Schwartz foresaw very clearly, and their pioneering work [1] provided many of the results that were to be useful later.

This manifold influence of distributions on PDE theory is best exemplified, in my opinion, in the statement and the proof of an important theorem of Bernard Malgrange, and it is perhaps worthwhile to give here a detailed exposition (and proof!) of this result. The reader who would be willing to go carefully through all the steps of the reasoning which we now present will encounter, in succession and meaningfully "embodied," the main aspects of the distributions impact.

First of all, what is the problem we wish to solve? We deal with a linear PDE with constant coefficients, in n variables, which we write

$$Pu = f. \tag{1}$$

The right hand side f is given: it is a complex-valued function in some (given)

Professor Treves received his Sorbonne PhD under Laurent Schwartz in 1958. He held positions at Berkeley, Yeshiva, and the Sorbonne, and visiting posts at Rio de Janeiro, Brasilia, and the Tata Institute, and is now at Purdue University. He was a Sloan Fellow from 1960 to 1964. His extensive research in partial differential equations includes the books *Topological Vector Spaces, Distributions and Kernels*, Academic Press 1967, *Linear Partial Differential Equations with Constant Coefficients*, Gordon and Breach 1967, and *Locally Convex Spaces and Linear Partial Differential Equations*, Springer Grundlehren series, 1967. *Editor.*

open subset Ω of the n-dimensional Euclidean space R^n; moreover f is infinitely differentiable, that is, f belongs to the space $\mathcal{C}^\infty(\Omega)$. Beyond this, f is arbitrary. We want to know whether there exists a solution u to (1). But we also want this solution to be infinitely differentiable in the open set Ω. Malgrange's theorem will provide a necessary and sufficient condition, bearing on the open set Ω and on the differential operator P, insuring that given any f in $\mathcal{C}^\infty(\Omega)$, there is u in $\mathcal{C}^\infty(\Omega)$ satisfying (1).

The reader cannot fail to observe that the statement of the problem involves no distributions, in fact that the problem is most "classical" in nature. Still the only proof known today (and to be found below) makes extensive use of distribution theory and of its most revolutionary aspects.

We must state the necessary and sufficient condition for solvability of Eq. (1). Let us write explicitly the differential operator P:

$$P = P(\partial/\partial x) = \sum c_p(\partial/\partial x)^p,$$

where the sum is finite, the c_p are complex numbers, and we have used the multi-index notation, namely that if $p = (p_1, \cdots, p_n)$,

$$(\partial/\partial x)^p = (\partial/\partial x_1)^{p_1} \cdots (\partial/\partial x_n)^{p_n}.$$

We set then (writing $|p| = p_1 + \cdots + p_n$)

$$\check{P} = P(-\partial/\partial x) = \sum (-1)^{|p|} c_p(\partial/\partial x)^p,$$

observing that if ϕ and ψ are two $\mathcal{C}^\infty$ functions with compact support in the whole space R^n, integration by parts yields

$$\int \phi(x)P(\partial/\partial x)\psi(x)dx = \int P(-\partial/\partial x)\phi(x)\psi(x)dx. \tag{2}$$

DEFINITION 1. *The open set Ω is said to be P-convex if to every compact subset K of Ω there is another compact subset K' of Ω such that, given any $\mathcal{C}^\infty$ function ϕ with compact support in Ω, the following is true*:

$$\textit{if } \check{P}\phi = 0 \textit{ outside of } K \textit{ then } \phi = 0 \textit{ outside of } K'. \tag{3}$$

Before we state the theorem, let us note that partial differentiations, and therefore any operator such as P, define *linear mappings* of $\mathcal{C}^\infty(\Omega)$ into itself. And solvability of Eq. (1) in the sense we have given to it simply means that the linear map

$$P: \mathcal{C}^\infty(\Omega) \to \mathcal{C}^\infty(\Omega)$$

is *onto* (or *surjective*), i.e., that $P\mathcal{C}^\infty(\Omega) = \mathcal{C}^\infty(\Omega)$.

THEOREM (B. Malgrange, 1954). *We have $P\mathcal{C}^\infty(\Omega) = \mathcal{C}^\infty(\Omega)$ if and only if Ω is P-convex.*

Of course, such a result raises the question as to when is an open set P-con-

vex. Although this is a difficult question (no complete characterization has been found yet) we have some information. For instance, every convex set, in particular the whole space R^n, is $\boldsymbol{P}$-convex regardless of what the operator $\boldsymbol{P}$ may be (barring the case where all the coefficients c_p are zero!). If the operator $\boldsymbol{P}$ is elliptic, every open set is $\boldsymbol{P}$-convex. When the number of independent variables n is equal to two, the $\boldsymbol{P}$-convex sets are known for every operator $\boldsymbol{P}$ (see Hörmander [1], Th. 3.7.2).

Proof of the "only if" part: We assume that Eq. (1) always possesses a $\mathcal{C}^\infty$ solution u in Ω, for arbitrary $\mathcal{C}^\infty$ right-hand sides f. We must derive from this that Ω is $\boldsymbol{P}$-convex.

Let K be a compact subset of Ω and let $\Phi(K)$ denote the set of all the $\mathcal{C}^\infty$ functions with compact support in Ω such that $\check{\boldsymbol{P}}\phi=0$ outside of K. Now, one of the first steps in applying distribution techniques is to equip the function spaces under consideration with a locally convex topology, and this is always done by means of appropriate seminorms. First we equip $\mathcal{C}^\infty(\Omega)$ with its standard $\mathcal{C}^\infty$ topology: functions ϕ_ν converge in $\mathcal{C}^\infty(\Omega)$ if they and every one of their derivatives converge uniformly on every compact subset of Ω. This convergence can be defined by using the seminorms

$$|\phi|_{H,m} = \sup_{x \text{ in } H}\left(\sum_{|p|\leq m} |(\partial/\partial x)^p\phi(x)|\right),$$

where H is an arbitrary compact subset of Ω, and m any integer $\geqq 0$ (*sup* stands for the supremum, which here is a *maximum*). Thus ϕ converges to ϕ_0 in $\mathcal{C}^\infty(\Omega)$ if, given any H and any m, $|\phi-\phi_0|_{H,m}\to 0$. Provided with this topology, $\mathcal{C}^\infty(\Omega)$ is a locally convex topological vector space. Moreover, the topology can be defined by a metric, i.e., $\mathcal{C}^\infty(\Omega)$ is *metrizable* (we shall never use explicitly the metric), and it is easily seen to be *complete*: every Cauchy sequence converges in it. These two properties of being locally convex on one hand, and metrizable and complete, on the other, allow us to apply the main theorems of S. Banach: the Hahn-Banach theorem and the open mapping theorem.

Note that $\Phi(K)$ is a linear space; we equip it also with a topology: the one defined by the semi-norms

$$|P(-\partial/\partial x)\phi|_{K,m}, \qquad m = 0, 1, \cdots. \tag{4}$$

It is easy to see, by a variety of ways, that each seminorm (4) is in fact a norm, that is, if it vanishes, the function ϕ must be identically equal to zero. The best way (because it extends to the case where $\boldsymbol{P}$ has *variable* coefficients) is to use our assumption, namely that (1) is solvable: indeed, let us solve Eq. (1) with $f=\bar{\phi}$ (complex conjugate of ϕ). Then, by (2) we have:

$$\int \phi\bar{\phi}\, dx = \int \phi \boldsymbol{P} u\, dx = \int \check{\boldsymbol{P}}\phi u\, dx = 0 \quad \text{if } \check{\boldsymbol{P}}\phi = 0;$$

whence $\int|\phi|^2\, dx=0$ which implies $\phi\equiv 0$. Other methods use Hörmander's in-

equalities or the Paley-Wiener theorem (see below). At any rate, the linear space $\Phi(K)$ topologized by (4) is now metrizable and locally convex. However, it is *not* complete (in general).

We consider the bilinear functional

$$(\phi, f) \to \int \phi(x) f(x) dx$$

on the product space $\Phi(K) \times \mathcal{C}^\infty(\Omega)$. We are going to show that it is *separately* continuous (in each one of its arguments, ϕ and f). Fixing ϕ, the continuity with respect to f is trivial, since

$$\left| \int \phi f dx \right| \leqq |f|_{H,0} \int |\phi| \, dx,$$

if ϕ vanishes outside of the compact subset H of Ω. Next we keep f fixed and study the continuity with respect to ϕ. At this critical stage we use the fact that (1) has a solution u in $\mathcal{C}^\infty(\Omega)$ (it would suffice to assume that (1) has a distribution solution, and this fact would show that the solvability of Eq. (1) in the space of distributions, for arbitrary $\mathcal{C}^\infty$ right hand sides, implies its solvability in the space $\mathcal{C}^\infty(\Omega)$). We have

$$\int \phi f \, dx = \int \phi P u \, dx = \int \check{P} \phi u \, dx,$$

hence

$$\left| \int \phi f \, dx \right| \leqq |\check{P}\phi|_{K,0} \int_K |u| \, dx,$$

which implies the desired continuity.

We use now a general theorem of Dieudonné-Schwartz which states that *if E and F are two locally convex metrizable spaces and if moreover F is complete, every separately continuous bilinear functional on $E \times F$ is in fact continuous.* Thus our functional $\int \phi f dx$ must be continuous, which can be expressed by the fact that there is a compact set $K' \subset \Omega$, two integers m and m' and a constant $C > 0$ such that, for all ϕ in $\Phi(K)$ and all f in $\mathcal{C}^\infty(\Omega)$,

$$\left| \int \phi f \check{P} dx \right| \leqq C |\check{P}\phi|_{K,m} |f|_{K',m'}.$$

But this implies that $\int \phi f dx = 0$ whenever f vanishes identically in a neighborhood of K'. This is only possible if ϕ vanishes outside of K'.

Proof of the "if" part: Now we assume that Ω is P-convex and prove that Eq. (1) is always solvable in $\mathcal{C}^\infty(\Omega)$, i.e., that the mapping P acting on $\mathcal{C}^\infty(\Omega)$ is onto. If $\mathcal{C}^\infty(\Omega)$ were finite dimensional, which it is not!, a simple criterion of surjectivity would be that the transpose of P, ${}^tP$, acting on the dual of $\mathcal{C}^\infty(\Omega)$ (dual which is customarily denoted by $\mathcal{E}'(\Omega)$), is *one-to-one* (or *injective*). The

injectivity of the transpose is not any more enough to insure the surjectivity of the mapping when the space acted upon is infinitely dimensional. To palliate this, we turn to the topologized space $\mathcal{C}^\infty(\Omega)$ and apply a criterion due to S. Banach and valid for metrizable complete locally convex spaces—which is exactly what $\mathcal{C}^\infty(\Omega)$ is. The criterion concerns, here also, the transpose of the mapping under consideration. Let us state it in full generality:

LEMMA 1 (S. Banach). *Let E, F be two Fréchet spaces (i.e., locally convex, metrizable, and complete), T: $E \to F$, a continuous linear map, E', F' the respective duals of E and F. Then the two following properties are equivalent*:

(a) *T is onto*;

(b) *the transpose ${}^tT$ of T is one-to-one and its range is weakly closed.*

This statement (which is classical) calls for some comments. First of all by the *dual* of E we mean the linear space of *continuous* linear functionals on E, i.e., of continuous linear mappings $E \to C$ (the complex field). Second, by the transpose of T we mean the unique linear mapping

$$ {}^tT: F' \to E' $$

such that, given any x in E and any y' in F',

$$ \langle {}^tT(y'), x \rangle = \langle y', T(x) \rangle, $$

where $\langle\ ,\ \rangle$ is the bracket of the duality, either between E and E' or between F and F'.

Finally, what do we mean by weakly closed? The range of ${}^tT$ is a linear subspace of E' and it should be closed for the weak topology on E', which is simply *the topology of pointwise convergence* for (linear) functions defined in E.

We want to apply Lemma 1 in our present situation, which is to say when $E = F = \mathcal{C}^\infty(\Omega)$ and when $T = \boldsymbol{P}$. Therefore we have to check that (b) holds in this situation. In order to do this, we use the fact that $\mathcal{E}'(\Omega)$, the dual of $\mathcal{C}^\infty(\Omega)$, can be regarded as a space of distributions in Ω. We recall that the space of all distributions in Ω, $\mathcal{D}'(\Omega)$, is the dual of the space of test-functions, i.e., of $\mathcal{C}^\infty$ functions *with compact support* in Ω, $\mathcal{C}_c^\infty(\Omega)$ (suitably topologized). Now the natural injection

$$ \mathcal{C}_c^\infty(\Omega) \to \mathcal{C}^\infty(\Omega) $$

is continuous and has a dense image; therefore its transpose is a continuous linear mapping (say, for the weak topologies)

$$ \mathcal{E}'(\Omega) \to \mathcal{D}'(\Omega) $$

which is injective. We may then identify $\mathcal{E}'(\Omega)$ with its image, a linear subspace of $\mathcal{D}'(\Omega)$. Can we characterize this subspace of $\mathcal{D}'(\Omega)$? Yes, and the answer is extremely simple: *a distribution belongs to $\mathcal{E}'(\Omega)$ if and only if it has compact support* (this is seen at once by observing that a linear functional λ on $\mathcal{C}_c^\infty(\Omega)$, if it is to be continuous for the topology induced by $\mathcal{C}^\infty(\Omega)$, must satisfy for some

constant $C>0$, some compact subset K of Ω, and some integer m,

$$|\lambda(\phi)| \leqq C|\phi|_{K,m} \quad \text{for all } \phi \text{ in } \mathcal{C}_c^\infty(\Omega);$$

therefore $\lambda(\phi)=0$ whenever ϕ vanishes in a neighborhood of K, i.e., the *support* of λ is contained in K).

We may already check the first part of Condition (b), that ${}^t\boldsymbol{P}$ acting on distributions with compact support, is one-to-one. Indeed, we apply Fourier transformation on the equation

$${}^t\boldsymbol{P}\lambda = 0;$$

this yields at once

$$P(-i\xi)\hat{\lambda}(\xi) = 0,$$

but the Paley-Wiener-Schwartz theorem states that the Fourier transform of a distribution with compact support is an entire function of exponential type; the product of such a function with a nonidentically zero polynomial cannot vanish identically unless the function itself does. But then Fourier transformation acting on distributions with compact support (or, for that matter, on tempered distributions) is one-to-one; therefore λ itself must be zero. This provides another proof of the fact that the seminorms (4) are norms.

As a last step, we must prove that the range of ${}^t\boldsymbol{P}$ is weakly closed. For this we use another criterion, due also to Banach:

LEMMA 2 (S. Banach). *Let E be a Fréchet space, $\{\mathcal{P}_\alpha\}$ a collection of seminorms on E such that the sets*

$$U_\alpha = \{e \text{ in } E;\ \mathcal{P}_\alpha(e) \leqq 1\}$$

form a basis of neighborhoods of zero in E. Let U_α^0 denote the polar of U_α, that is the subset of E' consisting of those linear functionals e' such that

$$|\langle e', e\rangle| \leqq 1 \quad \text{for all } e \text{ in } U_\alpha.$$

A linear subspace M' of E' is weakly closed if and only if its intersection with every polar set U_α^0 is weakly closed.

We apply this, taking $\mathcal{P}_\alpha$ to be the seminorm

$$N|\phi|_{K,m}, \qquad N, m = 1, 2, \cdots,$$

K ranging over all the compact subsets of Ω. Then the assumption in Lemma 2 is satisfied. *The polar U_α^0 consists exclusively of distributions having their support contained in K.* Indeed, saying that λ belongs to U_α^0 is equivalent with saying that for all functions ϕ in $\mathcal{C}_c^\infty(\Omega)$,

$$|\langle\lambda, \phi\rangle| \leqq N|\phi|_{K,m};$$

therefore $\langle\lambda, \phi\rangle=0$ whenever ϕ vanishes in a neighborhood of K, which is simply saying that the support of λ is contained in K.

We have to prove that

$$^{t}P\mathcal{E}'(\Omega) \cap U_\alpha^0 \tag{5}$$

is weakly closed. Observe that U_α^0 is always weakly closed (polar sets always are). Let $\{\lambda_j\}$ be a net, or a filter in the intersection (5) converging weakly to $\lambda_0 \in \mathcal{E}'(\Omega)$. Since λ_0 belongs to U_α^0 all we have to do is to show that λ_0 belongs to the range of ^{t}P, i.e., that there is μ_0 in $\mathcal{E}'(\Omega)$ such that

$$\lambda_0 = {}^{t}P\mu_0.$$

Now, each λ_j is of the form

$$\lambda_j = {}^{t}P\mu_j, \qquad \mu_j \text{ in } \mathcal{E}'(\Omega). \tag{6}$$

We use now a basic theorem in the theory of PDEs with constant coefficients:

THEOREM (B. Malgrange, L. Ehrenpreis, 1952). *Every linear PDE with constant coefficients (nonidentically zero) has a fundamental solution.*

We apply this result to $\check{P} = P(-\partial/\partial x)$, acting on distributions. But the definition of this action, which generalizes the integration by parts formula (2), shows that $\check{P} = {}^{t}P$. We see thus that there is a distribution E in R^n such that

$$^{t}PE = P(-\partial/\partial x)E = \delta, \quad \text{the Dirac distribution.}$$

Next we use convolution, in relation to (6); we convolve both sides with E. But in order to differentiate a convolution of two distributions (one of which, at least, must have compact support: this is the case here), we may differentiate indifferently anyone of the factors. Therefore:

$$E * \lambda_j = E * P(-\partial/\partial x)\mu_j = \{P(-\partial/\partial x)E\} * \mu_j = \delta * \mu_j = \mu_j.$$

We exploit the fact that $\lambda \to E * \lambda$ is a continuous linear map of $\mathcal{E}'(R^n)$, the distributions with compact support in R^n, into $\mathcal{D}'(R^n)$, the space of all distributions in R^n (this type of continuity is one of the standard features of operations on distributions; all duals carry their weak topologies). We obtain that the $\mu_j = E * \lambda_j$ converge weakly to a distribution μ_0 in R^n. We must have, by continuity,

$$\mu_0 = E * \lambda_0.$$

By applying $P(-\partial/\partial x)$ to both members, we obtain

$$\lambda_0 = {}^{t}P\mu_0.$$

Thus λ_0 belongs to the range of ^{t}P but not as a mapping in $\mathcal{E}'(\Omega)$, since we do not know yet whether λ_0 has a compact support contained in Ω. This is what remains to be shown.

Suppose we knew that all the μ_j had their support in a fixed compact subset K' of Ω. Then this would also be true of μ_0 since such a property subsists when we take weak limits. What we know is that $P(-\partial/\partial x)\mu_j$ has its support contained in K (see remark following Lemma 2). At this very last stage of the

proof, we use the P-convexity of Ω. Recalling that ${}^tP$ extends $\check{P}$, we would have the desired property if (3), in Definition 1, were valid for distributions with compact support, in addition to being valid for test functions. The proof will be completed if we show this to be the case.

Let ϕ_k, $k=1, 2, \cdots$, be a sequence of test functions in R^n converging to the Dirac distribution δ in standard fashion: for instance, $\phi_k \geqq 0$; ϕ_k vanishes identically outside of the sphere of radius $1/k$; $\int \phi_k dx = 1$ for all k. Consider the convolution $\phi_k * \mu$, where μ in $\mathcal{E}'(\Omega)$ is such that the support of $P(-\partial/\partial x)\mu$ is contained in K. As soon as k is large enough so that the neighborhood of order $1/k$ of the support of μ is a compact subset of Ω, we have that $\phi_k * \mu$ belongs to $\mathcal{C}_c^\infty(\Omega)$. Moreover,

$$P(-\partial/\partial x)(\phi_k * \mu) = \phi_k * P(-\partial/\partial x)\mu$$

vanishes outside the neighborhood of order $1/k$ of K. By Property (3) we derive that there is a compact subset K' of Ω such that

$$\phi_k * \mu \equiv 0 \quad \text{outside of } K',$$

as soon as k is large enough. By taking k to the limit $+\infty$, we conclude that μ vanishes outside of K', which is what we wanted to prove.

An invited address before the Mathematical Association of America in New Orleans, January 1969.

Bibliographical References

N. Bourbaki, [1]. Espaces Vectoriels Topologiques, Act. Scient. et Industr., no. 1189, 1229, Paris 1953, 1955.

J. Dieudonné, and L. Schwartz, [1]. La dualité dans les espaces ($\mathfrak{F}$) et ($\mathfrak{LF}$), Ann. Inst. Fourier Grenoble, 1 (1949) 61-101.

L. Hörmander, [1]. Linear Partial Differential Operators, Springer, New York, 1963.

Further Reading

H. Bremermann, Distributions, Complex Variables and Fourier Transforms, Addison-Wesley, Reading, Mass., 1965.

R. W. Carroll, Abstract Methods in Partial Differential Equations, Harper Row, New York, 1969.

A. Friedman, Generalized Functions and Partial Differential Equations, Prentice-Hall, Englewood Cliffs, N.J., 1963.

I. M. Gelfand, and G. Silov, Generalized Functions, English translation, Academic Press, New York, 1964.

L. Hörmander, [1] (see above).

J. Horváth, Topological Vector Spaces and Distributions, Addison-Wesley, Reading, Mass., 1966.

G. Köthe, Topologische lineare Räume, Vol. I, Springer, New York, 1960.

L. Schwartz, Théorie des Distributions, Hermann, Paris, 3rd ed., 1966.

F. Treves, Linear Partial Differential Equations with Constant Coefficients, Gordon and Breach, New York, 1967.

———, Locally Convex Spaces and Linear Partial Differential Equations, Springer, New York, 1967.

———, Topological Vector Spaces, Distributions and Kernels, Academic Press, New York, 1967.

K. Yosida, Functional Analysis, Springer, New York, 1965.

CODING THEORY: A COUNTEREXAMPLE TO G. H. HARDY'S CONCEPTION OF APPLIED MATHEMATICS

NORMAN LEVINSON, Massachusetts Institute of Technology

1. Introduction. A major theme of G. H. Hardy in "A Mathematician's Apology" [5] is the division of mathematics into pure mathematics, "the 'real' mathematics of the 'real' mathematicians which is almost wholly 'useless' " [5, p. 119] and applied mathematics, which he regarded as dull and trivial. In contrast to the harmlessness and innocence of 'real' mathematics, the "trivial mathematics on the other hand has many applications in war." See [5, p. 141]. Hardy exults particularly in the uselessness of number theory which, if "real mathematics" were useful, could be exploited for evil as well as good. Hence "Gauss and lesser mathematicians may be justified in rejoicing that there is one science at any rate, and that their own, whose very remoteness from ordinary human activities should keep it gentle and clean." See [5, p. 121].

Hardy did approve of theoretical physics as exemplified by relativity and quantum mechanics, but regarded them as quite useless [5, p. 135]. If time has shown him wrong about this, it can be argued that in these subjects he was not an expert, and therefore the real test of his ideas concern pure mathematics.

Here we shall show how coding theory refutes Hardy's notion. Finite fields, also called Galois fields, and theorems from number theory play a central role in coding theory. In some areas of applied mathematics, the role of pure mathematics is often at best one of reassurance, such as in providing a nonconstructive existence theorem or a uniqueness theorem, but not in providing the computational or analytic procedures that yield the actual results. In practice the procedures used may involve more intuition and experience than rigor. This is not the case in coding theory, where pure mathematics supplies the constructive procedure for carrying out coding. This may surprise applied mathematicians more than it will pure mathematicians. To accommodate those applied readers, our account will not require familiarity with finite fields or number theory. Rather we shall start with the problem of error correction in the transmission of information by use of codes, and show how this leads to the introduction of a certain mathematical object which is in fact a finite field. Cyclotomic polynomials, a discovery of Gauss, will play a key role.

Quadratic residues and the law of quadratic reciprocity (which Hardy [5, p. 92] regarded as one of the most beautiful theorems of mathematics) also enter

Professor Levinson studied with Norbert Wiener and received the Sc.D. degree in 1935. He was a travelling Fellow at Cambridge University in 1934–35 and an NRC Fellow in 1935–37. He has been on the MIT staff since 1937, except for a year as Guggenheim Fellow at the Mathematics Institute, Copenhagen and a year at the University of Tel Aviv. His main research interests are transforms, entire functions, probability, and differential equations. He is the author of the AMS Colloquium volume, *Gap and Density Theorems*, and (with E. Coddington) *Ordinary Differential Equations*. Professor Levinson received the AMS Bôcher Prize and he is a member of the National Academy of Sciences. *Editor.*

coding theory [1, pp. 173, 354]. Another working tool is the Chinese remainder theorem [1, p. 339]. Hardy [5, p. 113] discusses the aesthetic quality of "real mathematics." Here the highly regarded theorems and their proofs possess "a very high degree of *unexpectedness*, combined with *inevitability* and *economy*." This is true of the manner in which finite fields enter coding theory, as we shall see.

Not unexpectedly, finite fields were introduced into coding theory mainly by men trained as mathematicians. Some of the early work was apparently not published. The particular development which will be described here, the BCH codes, is due independently to Bose and Ray-Chauduri [2] and to Hocquenhem [6]. What is most important for the actual usefulness of the method, an efficient decoding process for these codes was discovered by an engineer, Peterson [7]. The BCH codes were generalized considerably by Gorenstein and Zierler [3].

Berlekamp [1, p. vii] states that "the essential limitation of all coding and decoding schemes . . . (has been) the complexity (and *cost*) *of the decoder*. The important work of Reed and Solomon (1960), Bose and Chaudhuri (1960), Gorenstein and Zierler (1961), and Peterson (1961) marked the advent of a new approach to this problem. By associating each digit of certain codes with an element in a Galois field, it was found possible to derive an algebraic equation whose roots represent the locations of the channel errors. . . . As a consequence it is now possible to build algebraic decoders which are orders of magnitude simpler than any that have previously been considered."

The notation used below mainly conforms with that used in Berlekamp [1].

2. Coding. Here a *message* will mean a finite ordered sequence of two symbols which it is desired to transmit through a channel. For example the channel may be a cable or a radio frequency band. It will be convenient to designate the two symbols as 0 and 1. A sequence of k such symbols may be regarded as a *binary k-vector* $(a_1, a_2, \cdots, a_k)$, where each a_j is either 0 or 1. Clearly there are 2^k binary k-vectors. If the transmission channel is noisy, the received vector may differ from the one sent, that is, the transmission process may introduce errors. One way to improve reliability is to repeat the message several times. This is an example of the use of *redundancy*, that is the transmission of more than the k binary digits contained in the original message in order to improve the reliability of the transmission process.

Simple repetition is not efficient. In general, a binary n-vector is transmitted with $n=k+r$, where k is the number of binary digits which form a message and r is the number of redundant digits. These redundant digits are determined according to some rule by the k digits of the message. The process of constructing the redundant n-vector from the message k-vector is called *encoding*. While there are 2^n binary n-vectors, the encoding process leads to a subset of 2^k of these, which may be called *code-vectors*. Because of errors in transmission, the n-vectors which are received need not be code-vectors. The process of correcting the received n-vector and extracting the original k-vector

is called *decoding*. The arithmetic operations in encoding and decoding will be carried out modulo 2, that is, $1+1=0$. This is equivalent to binary addition with no carry-over, and hence is a process easy to design into an electronic computer. The binary arithmetic to be used here need involve only 0 and 1, with rules $0+0=0$, $0+1=1+0=1$, $1+1=0$, $0\cdot 0=0\cdot 1=1\cdot 0=0$ and $1\cdot 1=1$. With these rules, 0 and 1 form a field of two elements, which is known as GF(2) the Galois field of two elements. (This is *not* the place where finite fields play a crucial role in coding theory, since GF(2) by itself is rather trivial.) All arithmetic that follows involving vectors, matrices, and polynomials will be carried out in GF(2). We recall that modulo 2 all even integers may be replaced by 0 and all odd integers by 1.

3. Hamming single error correcting code. Suppose a channel is sufficiently reliable so we can assume that if a binary n-vector is transmitted, then the received binary n-vector contains an error in at most one entry. How much redundancy will allow the position of the error to be determined? Suppose m is a positive integer and set $n=2^m-1$. (This assumption about n here and later is more restrictive than necessary, but is sufficient to illustrate the basic ideas.) A binary number b that could designate which of the n received binary digits contains an error must itself have m digits, because it requires m binary digits to represent the positive integers not exceeding 2^m-1. The occurrence of no error can be designated by all m digits of b zero. As an example, let $m=4$ and hence $n=15$. Then the four-place binary numbers starting with 0001 and ending with 1111 represent all integers from 1 to 15. The above remarks suggest that it may be possible to correct a single error in the transmission of an n-vector, where $n=2^m-1$, if $r=m$ and hence $k=n-m$. A feasible method for doing so was discovered by Hamming [4]. Suppose again that $m=4$ so that $n=15$ and $r=4$; hence $k=11$. All vectors which will be considered from here on will be column vectors which (for typographical reasons) may also be written in row form. Denote the code vectors by $\mathbf{C}=(C_1, C_2, \cdots, C_{15})$, where the C_j are binary digits. Let H be a matrix of 15 columns, each column a binary vector with four entries, all columns distinct, and none identically zero:

$$H = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \cdots 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \quad\ \ 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \quad\ \ 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \cdots 1 \end{bmatrix}.$$

(If a column of H is written as a row, then that row considered as a binary number, represents the column number. This is convenient but not essential.)

Let the eleven entries $C_3, C_5, C_6, C_7, C_9, C_{10}, \cdots, C_{15}$ of $\mathbf{C}$ be the entries of the message k-vector. Determine C_1, C_2, C_4, and C_8 so that

$$HC = 0. \tag{3.1}$$

This is possible because the square matrix made up of the eighth, fourth, second,

and first columns of H is nonsingular. (Indeed it is the unit matrix.) By (3.1) all code vectors $\boldsymbol{C}$ are orthogonal to the rows of H. Since the arithmetic above is all modulo 2, C_1, C_2, C_4, and C_8 are of course binary digits. This determination of the code vector $\boldsymbol{C}$ completes the encoding process.

Suppose for the moment that errors occur in several of the digits in the transmission of $\boldsymbol{C}$ so that the received binary n-vector $\boldsymbol{R}$ does not coincide with $\boldsymbol{C}$. Performing addition mod 2 componentwise, let the binary n-vector $\boldsymbol{E}$ be defined by

$$\boldsymbol{E} = \boldsymbol{R} - \boldsymbol{C} = \boldsymbol{R} + \boldsymbol{C}. \tag{3.2}$$

If the jth entry of $\boldsymbol{R}$, namely R_j, and that of $\boldsymbol{C}$, namely C_j, coincide, then $E_j = 0$. But if $R_j \neq C_j$, then $E_j = 1$. Hence the vector $\boldsymbol{E}$ has entries differing from 0 at precisely those positions *where an error in transmission occurs.* Consider now

$$H\boldsymbol{R} = H\boldsymbol{C} + H\boldsymbol{E} = H\boldsymbol{E},$$

where use is made of (3.1). If no error has occurred in transmission, then $E = 0$, and so $H\boldsymbol{R} = 0$. If exactly one error has occurred in transmission and this error is in the jth term, then $E_j = 1$ and so

$$H\boldsymbol{R} = H^{(j)},$$

where $H^{(j)}$ is the jth column of H. Clearly knowledge of the vector $H^{(j)}$ determines j. (Indeed with the H above, writing $H^{(j)}$ in row form gives the binary representation of j.) Hence if at most one error occurs, $H\boldsymbol{R}$ determines at which, if any, entry the error occurs. Thus the binary n-vector $\boldsymbol{E}$ is determined, and since $\boldsymbol{C} = \boldsymbol{R} + \boldsymbol{E}$, the vector $\boldsymbol{C}$ is now available. If the entries C_1, C_2, C_4, and C_8 are discarded, the resulting binary k-vector is the original message. This process of reconstructing the message from $\boldsymbol{R}$ is the decoding process. Of course if more than one error occurs, that is, if $\boldsymbol{E}$ has two or more entries which are 1, the above procedure is not valid.

To correct more than one error in $\boldsymbol{R}$, the received binary n-vector, one would expect to increase r, the number of redundant digits, and hence to decrease k. Moreover, for $H\boldsymbol{R}$ then to yield the error locations in $\boldsymbol{R}$, one would expect H to have more rows. The decoding process can be expected to be least complicated if H has a structural pattern based on a reasonably simple mathematical algorithm. A comparatively simple mathematical scheme for locating, at least in principle, up to a prescribed number of errors is used in the BCH codes mentioned earlier. The BCH codes make use of finite fields.

4. The fields GF(2^m). (The reader familiar with finite fields can skim this section.) A column of an m-rowed matrix H may be regarded as a binary m-vector (where of course $m \geqq 1$). If x is an indeterminate and the a_j are all in GF(2), then the polynomial $\sum_0^{m-1} a_j x^j$ of degree $m-1$ can be used to represent a binary m-vector with entries a_j, where $0 \leqq j \leqq m-1$. (Here it is convenient to start the index j at 0.) Since each a_j, for $0 \leqq j \leqq m-1$, can be either 0 or 1, there are a total of 2^m of these polynomials of degree not exceeding $m-1$. Addi-

tion of these polynomials is equivalent to vector addition in GF(2) and leads again to one of the 2^m polynomials.

These rather trivial observations become profound if one further requires that the product of any two of the polynomials is again such a polynomial. Since the degree of the product of two polynomials is the sum of the degrees of each, the degree of the product will be at most $m-1$ only if some artifice is used. One way to achieve this is to compute polynomials modulo a fixed polynomial of degree m which we shall call $f(x)$. Thus if $P(x)$ is a polynomial, then $P(x)$ is equivalent to $P_1(x)$ where $P_1(x)$ is the remainder obtained in dividing $P(x)$ by $f(x)$; thus

$$P(x) = P_1(x) \pmod f$$

if

$$P(x) = J(x)f(x) + P_1(x),$$

where $J(x)$ is a polynomial and the degree of $P_1(x)$ is at most $m-1$. The arithmetic in the division of course is performed in GF(2). In particular $P(x)=0 \pmod f$ if and only if $f(x)$ is a divisor of $P(x)$.

As already stated there are exactly 2^m polynomials of degree not exceeding $m-1$ and with coefficients in GF(2). In this paragraph we shall exclude the null polynomial for which all $a_j=0$. Thus there remain $n=2^m-1$ polynomials. The manipulation of the m-vectors represented by these polynomials becomes particularly simple if the sequence

$$\{x^j\}, \qquad 0 \leqq j \leqq n-1, \pmod f \tag{4.1}$$

generates all n nonnull polynomials.

Example: $m=2$, $f(x)=x^2+x+1$; hence $n=3$. Then the sequence $1, x, x^2$, $\pmod f$ is $1, x, 1+x$, which are the three nonnull polynomials of degree not exceeding $m-1=1$ with coefficients in GF(2).

We shall show that the sequence (4.1) generates all n of the nonnull polynomials if

$$x^n = 1 \pmod f \quad \text{and} \quad x^k \not\equiv 1 \pmod f \qquad \text{for } 1 \leqq k < n. \tag{4.2}$$

(This is the statement that (4.1) should form a cyclic group of order n.) The x^j, $0 \leqq j \leqq n-1$, are distinct $\pmod f$. Indeed suppose that

$$x^j = x^k \pmod f, \qquad 0 \leqq j < k \leqq n-1.$$

Then multiply the above equation by x^{n-k} and use the first equation of (4.2) to get

$$x^{n-(k-j)} = 1 \pmod f.$$

Since $k>j$ this violates (4.2). Furthermore no $x^j=0 \pmod f$, since multiplying by x^{n-j}, we should have $1=0 \pmod f$ or $f(x)$ divides 1, which is impossible since degree $f=m\geqq 1$. Hence the sequence (4.1) of n elements are all distinct $\pmod f$

and none is the null element. Therefore the sequence (4.1) generates all n of the nonnull polynomials if (4.2) holds.

If $y=x^i$ for some fixed $i\geqq 1$, then the least positive integer λ for which $y^\lambda=1 \pmod f$ is called the *order* of y. If $y^k=1 \pmod f$ for some $k\geqq 1$, then k is a multiple of λ. Indeed let $k=q\lambda+s$ where $q\geqq 0$ and $0\leqq s<\lambda$. Then $1=y^k =y^{\lambda q+s}=y^s \pmod f$. From the definition of λ, since $s<\lambda$, this implies $s=0$ and proves the following special case of a classical result:

LEMMA 4.1. *Let y be a power of x and let y be of order λ. If $y^k\equiv 1 \pmod f$, then k is a multiple of λ.*

From (4.2), $f(x)$ must be a factor of x^n-1. Let us now again take the case $m=4$ (so that $n=15$) and enumerate certain particularly relevant factors of $x^{15}-1$. We revert to ordinary arithmetic, and note that if $x^3=1$ or if $x^5=1$, then certainly $x^{15}=1$. Hence x^3-1 and x^5-1 are factors of $x^{15}-1$. Of course $x-1$ is a factor of all of these. We now write the obvious identity

$$\begin{aligned} x^{15}-1 &= (x-1)\frac{x^3-1}{x-1}\,\frac{x^5-1}{x-1}\left(\frac{(x^{15}-1)(x-1)}{(x^3-1)(x^5-1)}\right) \\ &= Q^{(1)}(x)Q^{(3)}(x)Q^{(5)}(x)Q^{(15)}(x), \end{aligned} \tag{4.3}$$

where

$$Q^{(1)}(x) = x-1, \qquad Q^{(3)}(x) = x^2+x+1, \qquad Q^{(5)}(x) = x^4+x^3+x^2+x+1,$$

and (as can be verified) $Q^{(15)}(x)=x^8-x^7+x^5-x^4+x^3-x+1$.

REMARK: The polynomials $Q^{(j)}(x)$ above are examples of the cyclotomic polynomials of Gauss, Eisenstein, etc. A root of unity ρ is said to have order $j\geqq 1$ if j is the least exponent for which $\rho^j=1$. As in Lemma 4.1, the roots of $x^{15}-1$ must all have orders which are factors of 15. The above factorization (4.3) involves the roots of orders 1, 3, 5, and 15, and these occur precisely in $Q^{(1)}$, $Q^{(3)}$, $Q^{(5)}$, and $Q^{(15)}$ respectively. The polynomial $Q^{(j)}(x)$ has roots all of which are of order j.

REMARK: The choice $m=4$ is not entirely an accident. Note that the cases $m=3$ and $m=5$ would not serve nearly as well as illustrative examples because 7 and 31 are prime numbers, and so the analogue of (4.3) would be too simple to be revealing. The case $m=6$ becomes computationally rather long to serve as a suitable example.

Return again to GF(2); there $Q^{(1)}$, $Q^{(3)}$, and $Q^{(5)}$ remain the same in (4.3). But as can be readily verified,

$$Q^{(15)}(x) = (x^4+x^3+1)(x^4+x+1).$$

Now take $f(x)$ as one of the two quartic factors of $Q^{(15)}$, say

$$f(x) = x^4+x^3+1. \tag{4.4}$$

Then since $f(x)$ is a factor of $Q^{(15)}(x)$, it is a factor of $x^{15}-1$; therefore $x^{15}=1$

(mod f). By Lemma 4.1, the order of x must be a divisor of 15, so it is 1, 3, 5, or 15. But $f(x)$ is obviously not a divisor of $x-1$ or x^3-1 and can easily be shown not to be a divisor of x^5-1. Hence the order of x is not 1, 3, or 5, so it must be 15. (This is in fact a particular instance of an easily proved general property of the cyclotomic polynomials.) Therefore (4.2) is satisfied, so $1, x, x^2, \cdots, x^{14}$ (mod f) are the 15 cubic polynomials with coefficients in GF(2), none of which is the null polynomial. Given any x^j, $1 \leqq j \leqq 14$, then x^{15-j} is obviously its inverse (mod f). Thus these polynomials form a group under multiplication (mod f). (It is of course the cyclic group.) If the null polynomial is adjoined, then the 16 polynomials obviously form a group under addition. It follows readily that mod f these 16 cubic polynomials with coefficients in GF(2) form a field of 16 elements. This field is known as GF(16).

REMARK: The polynomial (4.4) is irreducible, that is, it cannot be written as the product of two lower degree polynomials in the arithmetic of GF(2). Indeed if it could, we should have

$$f_1(x)f_2(x) = f(x).$$

But $f_1(x)$ is a member of GF(16), hence has an inverse; the same is true of $f_2(x)$. If we multiply by these, we obtain $1=0$ (mod f), which is impossible.

REMARK: In principle the entire above procedure can be carried out to establish the existence of GF(2^m) for any m, and to specify an appropriate $f(x)$ of degree m. Actually to treat the general case it is necessary to develop a little more theory concerning $Q^{(i)}(x)$ and the irreducible polynomials with coefficients in GF(2), [1].

A more convenient way to indicate that we are working mod $f(x)$ is to let α denote a root of $f(x)$. Therefore $\alpha^4+\alpha^3+1=0$ and so any polynomial in α is automatically equivalent to a binary cubic in α. It is the cubic which one computes working mod f since the only property of α that is used is $f(\alpha)=0$. Thus the elements of GF(16) may be designated by the binary cubics in α.

Summary: Let α be a root of $\alpha^4+\alpha^3+1=0$. (Only this equation, and not the actual numerical value of α, is used.) Then each polynomial in α with coefficients in GF(2) is equal to a binary cubic polynomial in α. There are $2^4=16$ binary cubics. These form the field GF(16). Moreover $\{\alpha^j\}$, for $0 \leqq j \leqq 14$, generates the 15 nonnull binary cubics which together with the null polynomial make up GF(16). A binary cubic may be viewed as a binary 4-vector.

5. A multiple error correcting code. To show how finite fields enter into coding, let us continue with the case $m=4$, $2^m=16$. Suppose now it is desired to correct up to 3 errors in the transmission of the encoded vector $\boldsymbol{C}$ with $n=15$ entries. Since with $n=15$ the correction of one error required a 4-rowed matrix, as displayed above (3.1), it seems plausible to try to correct three errors with a 12-rowed matrix. This operating on $\boldsymbol{R}$ leads to a 12-vector which can be viewed as three 4-vectors and hence contains sufficient information to determine three integers between 1 and 15 and thereby locate up to three errors in $\boldsymbol{R}$. A systematic way to construct H is with its twelve rows arranged in three blocks of

four rows as follows:

$$H = \begin{bmatrix} 1 & \alpha & \alpha^2 & \alpha^3 & \cdots \alpha^{14} \\ 1 & \alpha^3 & \alpha^6 & \alpha^9 & \cdots \alpha^{42} \\ 1 & \alpha^5 & \alpha^{10} & \alpha^{15} & \cdots \alpha^{70} \end{bmatrix}.$$

Each power of α of course represents a binary 4-vector belonging to GF(16). Why the row blocks α^{2j} and α^{4j}, $0 \leqq j \leqq 14$, can be omitted will soon be apparent. (Since $\alpha^4 = \alpha^3 + 1$ in GF(2), the first block of four rows can be computed from $1\ \alpha\ \alpha^2 \cdots \alpha^{14}$ and is

$$\begin{matrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0. \end{matrix}$$

The second block of four rows of H, namely $1\ \alpha^3\ \alpha^6 \cdots \alpha^{42}$, consists of the first, fourth, seventh, tenth, and thirteenth columns of the above displayed matrix repeated three times. The third block of four rows consists of the first, sixth, and eleventh columns of the above displayed matrix repeated five times.)

The received binary 15-vector $\boldsymbol{R}$ with entries R_j, $0 \leqq j \leqq 14$, has the polynomial representation

$$R(x) = \sum_0^{14} R_j x^j.$$

(Here we are *not* computing $R(x) \bmod f$.) The application $H\boldsymbol{R}$ of H to $\boldsymbol{R}$ is the column vector with 12 entries

$$R(\alpha),\ R(\alpha^3),\ R(\alpha^5)$$

which is represented in row form as a triple of 4-vectors, each 4-vector being an element of GF(16).

As in (3.2), let $\boldsymbol{R} = \boldsymbol{C} + \boldsymbol{E}$. If $\boldsymbol{E} = 0$, then it is desirable that $H\boldsymbol{R} = 0$ and therefore we should have $H\boldsymbol{C} = 0$. Let

$$C(x) = \sum_0^{14} C_j x^j.$$

In terms of α, $H\boldsymbol{C}$ is the binary 12-vector $C(\alpha)$, $C(\alpha^3)$, $C(\alpha^5)$. For this 12-vector to be zero, the polynomial $C(x)$ of degree 14 should vanish for $x = \alpha$, α^3, and α^5. To make $C(x)$ vanish for $x = \alpha$, we can require that $f(x)$ be a factor of $C(x)$ in GF(2). It will be convenient now to designate $f(x)$ by $M_1(x)$. To find a polynomial, say $M_3(x)$, which has α^3 as a root, note that α^3 has order 5 and hence will be a root of $Q^{(5)}(x)$ of (4.3), where

$$Q^{(5)}(x) = 1 + x + x^2 + x^3 + x^4.$$

Denote $Q^{(5)}(x)$ by $M_3(x)$. Similarly α^5 has order 3 and hence is a root of $Q^{(3)}(x) = 1+x+x^2$, which we shall denote by $M_5(x)$. (The polynomials M_1, M_3, and M_5 are all minimum polynomials in the sense that no polynomials of lower degree with coefficients in GF(2) have α, α^3, and α^5 respectively as roots.) Let

$$g(x) = M_1(x)M_3(x)M_5(x).$$

Then the degree of $g(x)$ is 10, since M_1 and M_3 are of a degree 4 and M_5 of degree 2. Moreover $g(x)$ vanishes for $x=\alpha$, α^3, and α^5 because $M_1(\alpha)$, $M_3(\alpha^3)$, and $M_5(\alpha^5)$ are each zero. We now require that $g(x)$ be a factor of $C(x)$ to assure that $C(x)$ has α, α^3, and α^5 as roots. Recall that $\mathbf{C}$ is a vector with 15 entries. Let C_{10}, C_{11}, C_{12}, C_{13}, C_{14} be a message vector of $k=5$ binary digits. Choose $\sum_0^9 C_j x^j$ as the negative of the remainder of the quotient

$$\frac{C_{14}x^{14} + C_{13}x^{13} + \cdots + C_{10}x^{10}}{g(x)}$$

so $C(x)$ will indeed have $g(x)$ as a factor and hence α, α^3, and α^5 as roots. The above arithmetic is of course in GF(2). This is the encoding process with $n=15$, $k=5$, and $r=10$. The binary polynomial $g(x)$ is known as the *generator polynomial* of the code. Indeed a code-vector $\mathbf{C}$ is characterized by the fact that $C(x)$ is divisible by $g(x)$.

It will now be shown in principle at least that the 12-vector HR can be used to correct up to a maximum of 3 errors in transmission. Since $HC=0$, therefore $HR=HE$, and the 12 vector HE regarded as a triple of 4-vectors determines $E(\alpha)$, $E(\alpha^3)$, and $E(\alpha^5)$. We recall that the entries of $\boldsymbol{E}$ are 1 where an error occurs and 0 otherwise. Suppose 3 errors occur say at the entries i_1, i_2, and i_3 of $\boldsymbol{E}$. Then

$$\begin{aligned} E(\alpha) &= \alpha^{i_1} + \alpha^{i_2} + \alpha^{i_3}, \\ E(\alpha^3) &= \alpha^{3i_1} + \alpha^{3i_2} + \alpha^{3i_3}, \\ E(\alpha^5) &= \alpha^{5i_1} + \alpha^{5i_2} + \alpha^{5i_3}. \end{aligned}$$

It will be convenient to note that with the a_j in GF(2),

$$(5.1) \qquad \left(\sum a_j\alpha^j\right)^2 = \sum a_j^2\alpha^{2j} = \sum a_j\alpha^{2j},$$

because all cross products have $2=1+1$ as a factor. Suppose now that 3 errors occur at the positions i_4, i_5, and i_6, all distinct from i_1, i_2, and i_3 above, and suppose these errors lead to the same values for $E(\alpha)$, $E(\alpha^3)$, and $E(\alpha^5)$ as do i_1, i_2, and i_3. Then this leads to

$$\alpha^{ji_1} + \alpha^{ji_2} + \alpha^{ji_3} = \alpha^{ji_4} + \alpha^{ji_5} + \alpha^{ji_6}, \qquad j = 1, 3, 5$$

or

$$(5.2) \qquad \sum_{d=1}^{6} \alpha^{ji_d} = 0 \qquad j = 1, 3, 5.$$

But now applying (5.1) to the case $j=1$, (5.2) holds for $j=2$. Applying (5.1) to the case $j=2$ and then to $j=3$ gives (5.2) for $j=4$ and $j=6$. (That is why we omitted the even powers of α from the rows of H.) Thus

$$\sum_{d=1}^{6} \alpha^{j i_d} = 0 \qquad j = 1, 2, 3, 4, 5, 6. \tag{5.3}$$

The determinant of the above system is a Vandermonde determinant equal to

$$\alpha^{i_1+i_2+\cdots+i_6} \prod_{6 \geqq d > e \geqq 1} (\alpha^{i_d} - \alpha^{i_e}). \tag{5.4}$$

Each factor of (5.4) is

$$\alpha^{i_d} - \alpha^{i_e} = \alpha^{i_e}(\alpha^{i_d - i_e} - 1).$$

Since $0 \leqq i_d \leqq 14$, it follows that $0 < |i_d - i_e| \leqq 14$. Hence, since the order of α is 15, no factor of (5.4) is zero so the determinant is not zero. Thus the homogeneous system (5.3) is impossible. If the other cases—such as two sets of three errors but with some in common, or where one set or both sets have less than three errors—are considered, there are now fewer columns in the analogue to (5.3); hence some rows may be discarded leading again to a Vandermonde situation. Thus if there are at most three errors, then HR determines their locations uniquely.

Of course for successful decoding, the above uniqueness result, while reassuring, must be replaced with a reasonably simple constructive procedure for determining which entries of the vector $\boldsymbol{E}$, if any, are 1. It is the simple orderly structure of H in terms of powers of α that makes the mechanization of such a decoding procedure feasible [1, Chap. 7].

The BCH codes, based on conceptions from pure mathematics, are not unique in using unexpected parts of pure mathematics for coding. Among the codes there are for example euclidean-geometry codes [1, p. 375], projective-geometry codes [1, p. 376], tensor product codes [1, p. 346], and quadratic residue codes [1, p. 354].

The preparation of this paper was supported in part by the Office of Naval Research and by the National Science Foundation NSF GP-13778.

References

1. Elwyn R. Berlekamp, Algebraic Coding Theory, McGraw-Hill, New York, 1968.

2. R. C. Bose and D. K. Ray-Chaudhuri, On a class of error correcting binary group codes, Information and Control, 3 (1960) 68–79 and 279–290.

3. D. C. Gorenstein and N. Zierler, A class of error-correcting codes in p^m symbols, J. Soc. Indust. Appl. Math., 9 (1961) 207–214.

4. R. W. Hamming, Error detecting and error correcting codes, Bell System Tech. J., 29 (1950) 147–160.

5. G. H. Hardy, A Mathematician's Apology, Cambridge University Press, 1967 Edition.

6. A. Hocquenhem, Codes correcteurs d'erreurs, Chiffres, 2 (1959) 147–156.

7. W. W. Peterson, Error-correcting Codes, M.I.T. Press, 1961.

8. I. A. Reed and G. Solomon, Polynomial codes over certain finite fields, J. Soc. Indust. Appl. Math., 8 (1960) 300–304.

POSITIVE DEFINITE MATRICES

C. R. JOHNSON, Northwestern University

Historically, positive definite matrices arise quite naturally in the study of n-ary quadratic forms and assume both theoretic and computational importance in a wide variety of applications. For example, they are employed in certain optimization algorithms in mathematical programming, in testing for the strict convexity of scalar-valued vector functions (here, positive definiteness of the Hessian provides a sufficiency check), and are of basic theoretic importance in construction of the various linear regression models. These are only a few of the specific applications which may be added to the abstract interest of such matrices. We now concentrate specifically on the properties of the matrices themselves.

DEFINITION: *An $n\times n$ real matrix A, where n is a positive integer, is called positive definite if* $(\mathbf{x}, A\mathbf{x}) = \mathbf{x}^T A\mathbf{x} > 0$ *for all nonzero column vectors* $\mathbf{x}$ *in Euclidean n-dimensional space.*

We shall designate the set of all such matrices (which forms a subset of all $n\times n$ matrices) as Π_n.

Classically, it is customary to require also symmetry in the definition of positive definite, and we shall often concentrate on that proper subset of Π_n which consists of only the symmetric members of Π_n. We shall designate this subset as Σ_n, but at times we shall allow ourself to consider the more general case.

It will become clear that the classical concentration on Σ_n is convenient since it is much richer in algebraic properties, but also, from the standpoint of testing arbitrary matrices, it suffices to consider the theory of Σ_n. As is well known, any square matrix A can be written as the sum of a symmetric and a skew-symmetric matrix, $A = B + C$, where $B = (A + A^T)/2$ and $C = (A - A^T)/2$. We call B the *symmetric* and C the *skew-symmetric* part of A, in an unambiguous manner.

(1) REMARK. An $n\times n$ matrix A is positive definite if and only if the symmetric part of A is positive definite. (Actually, we show that any quadratic form is equivalent to a symmetric quadratic form.)

Proof. If C is the skew-symmetric part of A, then $(\mathbf{x}, C\mathbf{x}) = (C^T\mathbf{x}, \mathbf{x}) = (-C\mathbf{x}, \mathbf{x}) = -(C\mathbf{x}, \mathbf{x})$ implies $(\mathbf{x}, C\mathbf{x}) = 0$. Therefore, $(\mathbf{x}, A\mathbf{x}) = (\mathbf{x}, (B+C)\mathbf{x}) = (\mathbf{x}, B\mathbf{x})$, where B is the symmetric part of A. Thus, in a certain sense (e.g., that of testing for positive definiteness), it suffices to study the subset Σ_n. The possibility of generalizing some of the results to Π_n will also be discussed.

We first characterize positive definiteness (Σ_n) in terms of a basic matrix invariant.

This is the winning essay in the Contest for Undergraduates announced in the January 1969 MONTHLY. Mr. Johnson was a student at Northwestern University then, and now he is a graduate student at Cal Tech. He is a member of the MAA so his prize was changed from membership and a book to two books: William Feller, Introduction to Probability Theory and its Applications, 2 vols., John Wiley & Sons, 1968. The prize was donated by the publisher. *Editor.*

(2) THEOREM. *If A is $n\times n$ symmetric, then $A\in\Sigma_n$ if and only if all eigenvalues of A are positive.*

Proof. We first note that since A is symmetric, A may be diagonalized by some *orthogonal* matrix B, $B^TAB=D$ where $B^T=B^{-1}$ and D is diagonal with the necessarily real eigenvalues of A (Principal Axis Theorem, see e.g., Zelinsky, [3]). Let the eigenvalues of A be $\lambda_1, \cdots, \lambda_n$.

If all $\lambda_i>0$, let $\mathbf{y}=B^T\mathbf{x}$ so that $\mathbf{x}=B\mathbf{y}$. Then $\mathbf{x}^TA\mathbf{x}=(B\mathbf{y})^TA(B\mathbf{y})=\mathbf{y}^T(B^TAB)\mathbf{y}>0$ if $\mathbf{x}\neq 0$.

If A is positive definite, let $\mathbf{v}_i$ be an eigenvector corresponding to λ_i normalized so that $(\mathbf{v}_i, \mathbf{v}_i)=1$. Then $0<\mathbf{v}_i^TA\mathbf{v}_i=(\lambda_i\mathbf{v}_i, \mathbf{v}_i)=\lambda_i$ and all $\lambda_i>0$.

(3) COROLLARY. *If $A\in\Sigma_n$, then det $(A)>0$.*

Proof: $\det(A)=\lambda_1\cdots\lambda_n>0$, by (2).

To facilitate the following comments, we let $S\subset\{1, 2, \cdots, n\}$, properly, be an index set. Then, let A_S, where A is an $n\times n$ matrix, be the matrix obtained from A by eliminating the rows and columns indicated by S, thus reducing the size of A.

(4) THEOREM. *If $A\in\Pi_n$, then A_S is positive definite for any S. In particular, the diagonal elements of A are >0.*

Proof: Let $\mathbf{x}\neq 0$ be an n-vector with zeros as the components indicated by S and arbitrary components elsewhere. If $\mathbf{x}_S$ is the vector obtained from $\mathbf{x}$ by eliminating the (zero) components indicated by S, then $\mathbf{x}_S^TA_S\mathbf{x}_S=\mathbf{x}^TA\mathbf{x}>0$. Since $\mathbf{x}_S\neq 0$ is arbitrary, A_S is positive definite.

As a special case, we may let $S=\{1, \cdots, i-1, i+1, \cdots, n\}$ to show that the ith diagonal element of A is >0.

We are now in a position to characterize positive definiteness in another manner which may be viewed as a test, the familiar determinant criteria. For this we employ the following abbreviations. If S is of the form $\{i+1, i+2, \cdots, n\}$, we denote A_S as A_i, i.e., A_i is the $i\times i$ matrix formed from the "intersection" of the first i rows and columns of A.

(5) THEOREM. *If A is $n\times n$ symmetric, then $A\in\Sigma_n$ if and only if* $\det(A_i)>0$ *for $i=1, \cdots, n$.*

Proof. If $A\in\Sigma_n$, then $A_i\in\Sigma_i$ by (4) and because A_i is symmetric. Therefore $\det(A_i)>0$ by (3).

Unfortunately, there are no thoroughly pleasing proofs of the converse proposition, but the Inclusion Principle for eigenvalues (Franklin, [1]) will aid in the proof. We note that if $\alpha_1\geqq\cdots\geqq\alpha_n$ are the n eigenvalues of A, and $\beta_1\geqq\cdots\geqq\beta_{n-1}$ are the $n-1$ eigenvalues of A_{n-1}, then $\alpha_1\geqq\beta_1\geqq\alpha_2\geqq\beta_2\geqq\cdots\geqq\alpha_{n-1}\geqq\beta_{n-1}\geqq\alpha_n$. Now, if $\det(A_i)>0$ for $i=1, \cdots, n$, we may inductively show $A\in\Sigma_n$. Since $\det(A_1)>0$, $A_1\in\Sigma_1$. If $A_k\in\Sigma_k$ for $k<n$, all eigenvalues of A_k are >0, and thus by the Inclusion Principle, all eigenvalues of A_{k+1} are

greater than 0, except perhaps the smallest. But, let $\alpha_1 \geqq \cdots \geqq \alpha_{k+1}$ be the eigenvalues of A_{k+1}; then $\alpha_{k+1} = \det(A_{k+1})/\alpha_1 \cdots \alpha_k$ as in the proof of (3), and α_{k+1}, the quotient of two positive reals, is positive. Thus, $A_{k+1} \in \Sigma_{k+1}$, and $A = A_n \in \Sigma_n$ by induction.

We choose to state here without full proof a result which applies in a much more general setting than Π_n, but which is important in applications of positive definite matrices.

(6) THEOREM. *If $A \in \Pi_n$, then A has a unique factorization $A = LR$ into triangular factors where L is lower triangular with 1's on the diagonal and R is upper triangular with nonzero diagonal elements.*

Comments. This is a consequence of the more general theorem (e.g., see [1], p. 204) that A has such a factorization if and only if $\det(A_i) \neq 0$ for $i = 1, \cdots, n$. Since $A \in \Pi_n$, each $A_i \in \Pi_i$ by (4). Thus $\det(A_i) \neq 0$, for otherwise A_i would be singular and not positive definite. Since L and R are obviously invertible (with easily computed inverses), this theorem is helpful, for instance, in exhibiting solutions $\mathbf{x}$ to $A\mathbf{x} = \mathbf{y}$.

We now have three rather concrete characterizations of Σ_n. With little difficulty we may add a fourth which is not quite so concrete.

(7) LEMMA. *If $A \in \Sigma_n$, then there exists an invertible matrix P such that $P^TAP = I$, the identity matrix. Also $PP^T = A^{-1}$ which exists.*

Proof: We know there exists an orthogonal B, such that $B^TAB = C$, where C is diagonal with the eigenvalues of A. If we let $D = +C^{-1/2}$, then D is well defined. Now let $P = BD$. Then, P is invertible since B and D are, and $P^TAP = D^TB^TABD = DCD = CD^2 = CC^{-1} = I$. Also, $P^TAPP^T = P^T$ and $APP^T = I$ since P^T is invertible. Therefore, $A^{-1} = PP^T$.

As an aside, (7) points out that any symmetric positive definite quadratic form $\mathbf{x}^TA\mathbf{x}$ (in fact, any positive definite quadratic form by (1)), is equal to $(\mathbf{y}, \mathbf{y})$ under a suitable change of coordinates.

(8) THEOREM. *$A \in \Sigma_n$ if and only if $\exists Q$ invertible such that $A = Q^TQ$.*

Proof. If $A = Q^TQ$ and $y = Qx$, then $\mathbf{x}^TA\mathbf{x} = \mathbf{x}^TQ^TQ\mathbf{x} = \mathbf{y}^T\mathbf{y} = (\mathbf{y}, \mathbf{y}) > 0$ if $\mathbf{x} \neq 0$ since Q is assumed invertible.

If $A \in \Sigma_n$, then $A = (P^T)^{-1}P^{-1}$, with P invertible by (7). Let $Q = P^{-1}$; then $Q^T = (P^{-1})^T = (P^T)^{-1}$.

We now mention a few less fundamental but still important results which deal with the positive definiteness of some functions of positive definite matrices.

(9) THEOREM. *The matrix A belongs to Σ_n if and only if B^TAB belongs to Σ_m for each $n \times m$ matrix B such that $B\mathbf{y} = 0$ implies $\mathbf{y} = 0$.*

Proof. If the condition is satisfied, let $m = n$ and choose $B = I$. Then $I^TAI = A \in \Sigma_n$.

Conversely, let $A \in \Sigma_n$ and suppose B satisfies the conditions of the theorem. Then $\mathbf{y}^T(B^TAB)\mathbf{y} = (B\mathbf{y})^TA(B\mathbf{y}) > 0$ unless $B\mathbf{y} = \mathbf{0}$, that is, $\mathbf{y} = \mathbf{0}$. Hence, B^TAB is symmetric and in Σ_m.

(10) THEOREM. *If $A \in \Sigma_n$, then*

(a) *$cA \in \Sigma_n$ for $c > 0$ any real scalar;*

(b) *$(A+B) \in \Sigma_n$ if $B \in \Sigma_n$;*

(c) *$A^m \in \Sigma_n$ for m any integer;*

(d) *an $A^{1/p}$ exists $\in \Sigma_n$ for p a positive integer (by $A^{1/p}$ we mean a matrix B such that $B^p = A$);*

(e) *an A^r exists $\in \Sigma_n$ for r any rational number.*

Proof. (a) cA is symmetric, and $\mathbf{x}^T(cA)\mathbf{x} = c(\mathbf{x}^TA\mathbf{x}) > 0$ if $c > 0$ and $\mathbf{x} \neq \mathbf{0}$.

(b) $\mathbf{x}^T(A+B)\mathbf{x} = \mathbf{x}^T(A\mathbf{x}+B\mathbf{x}) = \mathbf{x}^TA\mathbf{x}+\mathbf{x}^TB\mathbf{x} > 0$ if $\mathbf{x} \neq \mathbf{0}$. Also, $A+B = (A+B)^T$ is symmetric.

(c) $(A^m)^T = (A^T)^m = A^m$ is symmetric. If $m = 0$, $A^m = I \in \Sigma_n$. If $m > 0$ and $C^TAC = D$ is an orthogonal diagonalization of A, then $A = CDC^T$ and $A^m = (CDC^T)(CDC^T) \cdots (CDC^T) = CD^mC^T$. Thus A^m is diagonalizable to D^m, and, therefore, has all eigenvalues > 0. Therefore, $A^m \in \Sigma_n$. If $m = -1$, then A^m exists in Σ_n by (7) and (8). If $m < -1$, $-m > 0$, $A^m = (A^{-1})^{-m}$, and $A^m \in \Sigma_n$.

(d) As before, we may write $A = CDC^T$, where C is orthogonal and D diagonal with positive diagonal elements. Define $A^{1/p} = CD^{1/p}C^T$, where $D^{1/p}$ is diagonal with diagonal elements the positive real pth roots of the diagonal elements of D. Then $(A^{1/p})^p = A$, $A^{1/p}$ is symmetric and has n positive real eigenvalues and is thus positive definite.

(e) Follows from (c) and (d).

The theorem, (10), might give us hope that a nontrivial algebraic structure might be imposed on Σ_n (or, perhaps, Π_n) so that it could be characterized as one of the more familiar algebraic objects. Unfortunately, however, this does not seem to be the case.

We may not employ an additive group structure since inverses do not exist in Σ_n. If A is positive definite, not only is $-A$ not positive definite, but it has essentially the opposite properties of A (such a matrix is called *negative definite*). Also, there is no additive identity in Σ_n or Π_n. Of course, both Σ_n and Π_n are closed and associative under addition.

The situation is interestingly different but only a little more well behaved if we attempt a multiplicative group structure on Σ_n. We have the identity matrix $I \in \Sigma_n$ and inverses exist in Σ_n by (10). Also, the multiplication is associative. But Σ_n is not closed under matrix multiplication. Not only is the product of two members of Σ_n not symmetric if they do not commute, but it may not even be positive definite as the following example in Σ_2 shows. Let

$$A = \begin{pmatrix} 1 & 3 \\ 3 & 10 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 1 & -3 \\ -3 & 10 \end{pmatrix}.$$

Both A and B are in Σ_2 by (5). However,

$$AB = \begin{pmatrix} -8 & 27 \\ -27 & 91 \end{pmatrix} \notin \Pi_2$$

by (4). In fact, AB is neither positive nor negative definite.

About all we can say, then, is that Π_n forms a semigroup under matrix addition. Since the other common algebraic structures are fundamentally more complex than the group, they are also precluded.

Some generalizations may be made without great difficulty on the proofs exhibited thus far. For instance, in the "only if" part of (2) our hypothesis is unnecessarily strong. If A is $n \times n$, it need only have n real eigenvalues. The proof is easily seen to depend only on the positive definiteness of A and the existence of n real eigenvalues with associated eigenvectors. Thus we may validly formulate the amending statement:

(2.1) THEOREM. *If A is $n \times n$ and has n real eigenvalues, then $A \in \Pi_n$ implies all the eigenvalues of A are positive.*

An entirely analogous statement may be made about the corollary, (3).

We may not, however, drop the requirement that A is real diagonalizable from (2.1) since a matrix may be positive definite and have complex characteristic roots as the following example shows. Let

$$A = \begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix}.$$

Then $A \in \Pi_2$ by (1) since the symmetric part of A is

$$\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}.$$

However, the characteristic polynomial of A is $\lambda^2 - 4\lambda + 5$ with roots $2 \pm i$. It should be clear, though, that under no circumstances may a member of Π_n have a nonpositive real eigenvalue; for then we could choose a nonzero eigenvector to violate the definition of Π_n. Thus, to generalize slightly on (7), a member of Π_n must always have nonzero determinant and be invertible.

In (10) parts (a) and (b) may clearly be generalized to all of Π_n, but part (c) is not always valid in Π_n. For instance, if

$$A = \begin{pmatrix} 1 & -4 \\ 2 & 10 \end{pmatrix},$$

$A \in \Pi_2$ by (1) and (5). But

$$A^2 = \begin{pmatrix} -7 & -44 \\ 22 & 92 \end{pmatrix}$$

which is not positive definite by (4). Also, this example should caution us to note that a matrix may have all its eigenvalues real and positive but not be $\in$ any Π_n. A^2 has eigenvalues $\{81, 4\}$.

If we allow complex-entried matrices and then define Σ_n^H to be the *Hermitian* matrices which are positive definite, it is clear also that our results concerning Σ_n may be modified to remain valid in Σ_n^H. In addition we should note that our notion of Π_n has no consistent analog in the complex-entried matrices. For instance, a matrix $A \in \Pi_n$ with respect to real vectors may not even have $\mathbf{x}^*A\mathbf{x}$ real when complex vectors are allowed (* means "conjugate transpose").

We have thus far formulated a theory of positive definite matrices. It is clear that we may analogously define another (disjoint) set of matrices by replacing "$>$" with "$<$" in the definition of positive definite. Such matrices are usually termed *negative definite*, and, suggestively, we might designate this set as $-\Pi_n$ since $A \in \Pi_n$ if and only if $-A \in -\Pi_n$. This, of course, is the key to the development of a theory of negative definite matrices which would proceed analogously (allowing for the peculiarities of negative numbers).

Positive (negative) *semi-definite* matrices may be defined by allowing the possibility of equality in the definition of Π_n (or $-\Pi_n$). Their theory proceeds similarly, but modified by allowance for 0 eigenvalues.

In the positive definite case, we have succeeded in establishing four characterizations through theorems: by eigenvalues ((2)); by determinants ((5)); by triangular decomposition ((6)); and by Q^TQ decomposition ((8)). This, plus the additional properties commented on, is largely sufficient to both mathematically describe and usefully apply positive definite matrices.

References

1. J. N. Franklin, Matrix Theory, Prentice-Hall, Englewood Cliffs, N. J., 1968.
2. A. S. Goldberger, Econometric Theory, Wiley, New York, 1966.
3. D. Zelinsky, A First Course in Linear Algebra, Academic Press, New York, 1968.

LOGIC VERSUS PEDAGOGY

MORRIS KLINE, New York University

Humble thyself, impotent reason.
PASCAL

1. The Current Emphasis on Logical Structure. There is no question that mathematics is distinguished from all other bodies of human knowledge in that it insists on deductive proof from explicitly stated axioms as the indispensable condition for the acceptance of its conclusions. This requirement has indeed conferred power on mathematics, for deductive proof has strengthened the

Professor Kline received his NYU degree under R. G. Putnam. He was a research assistant at the Institute for Advanced Study for two years, a physicist with Signal Corps Engineering Labs for three years, and has been on the NYU staff since the war. He was Director of the Courant Inst. Division of Electromagnetic Research, and is presently Chairman of Undergraduate Mathematics.

structure. Moreover, the organization of mathematics into deductive systems has given coherence to its vast contents, and the axiomatization inherent in this type of organization has made clear precisely what is presupposed and hence where such systems are applicable. It has also suggested abstractions which embrace several structures as, for example, the theory of groups.

The deductive organization of mathematics is now a popular mode of presentation in all instruction from the fourth or fifth grade up. Deductive proof is the be-all and end-all of teaching. The enthusiasm for this mode of presentation is somewhat understandable because it was only about 75 years ago, after over 2500 years of struggle, that deductive organization and, through it, rigor were achieved. Mathematicians may be now giving vent in their textbooks to the satisfaction that Poincaré expressed in 1900 at the International Congress of Mathematicians when he gloated [**11**], "One may say today that absolute rigor has been attained." As an expression of enthusiasm this emphasis on deduction and rigor might indeed be excused. However, when authors are challenged as to the pedagogical wisdom of such presentations they now rejoin that this is the way to understand mathematics. In other words, the deductive approach is being defended as the pedagogical approach. Deductive organization and proof are advocated as the answer to all of the difficulties which students have had in learning mathematics and the open sesame to the subject. No longer will the memorization of techniques be necessary. Mathematics will now be accessible and understandable to almost all students.

2. The Intuitive Approach. Opposed to the deductive approach is the intuitive approach. Admittedly the nature of intuition is somewhat vague. It denotes some direct grasp of the idea, whether it be a concept or proof. There may be a special intuitive faculty distinct from the logical faculty that criticizes and reasons. Whether or not there is an intuitive faculty there are specific and explicit aids to the intuition which enable it to function. Primarily it seems to rely upon the senses, for, as Aristotle first put it, there is nothing in the intellect that was not first in the senses, except, Leibniz added, the intellect itself. Hence one of the useful devices is a picture. Consider exhibiting several triangles to inculcate the idea as opposed to the definition: the union of three non-collinear points and the line segments joining them. How much more readily is the notion of continuity grasped when presented as a curve which can be drawn with an uninterrupted motion of a pencil rather than by the $\epsilon-\delta$ definition.

He was a Fulbright Lecturer and Guggenheim Fellow, and has spent several leaves in Stanford and at the Technische Hochschule in Aachen.

He has published extensively in topology (his original subject), electromagnetic theory, history and philosophy of mathematics, and pedagogy. His books include *Mathematics in Western Culture* (Oxford 1953), *Mathematics and the Physical World* (Crowell 1959), *Mathematics, A Cultural Approach* (Addison-Wesley 1962), *Electromagnetic Theory and Geometric Optics* (with I. W. Kay, Wiley 1965), *Calculus: An Intuitive and Physical Approach*, 2 vols. (Wiley 1967), and *Mathematics for Liberal Arts* (Addison-Wesley 1967). He has edited several collections, and he holds five patents on antenna systems. *Editor.*

The intuition may be appealed to through physical arguments. The derivative as a velocity at an instant gives meaning to the concept, and the argument that a ball thrown up into the air must have zero velocity at its highest point suggests that the derivative must be zero at a maximum of a function.

We shall include in the intuitive approach what are often called heuristic arguments. Through experience with actual objects a child can learn that $3+4=4+3$. The generalization that $a+b=b+a$ is heuristic. Likewise the fact that if $y=x^n$, $dy/dx=nx^{n-1}$ is readily inferred from the special cases $n=2$ and $n=3$. Reasoning by analogy and even probabilistic arguments are heuristic.

Intuition is not static. Just as one's intuition about what to expect in human behavior improves with experience so does the mathematical intuition. The latter may indeed suggest, as it did to Leibniz, that the derivative of a product of two functions is the product of the derivatives. The conclusion should be tested, another heuristic measure, and of course will be found to be false. Deeper analysis will show that what holds for limits of functions does not hold for derivatives, and the intuition will be sharpened by this experience.

Clearly the intuitive approach can lead to error, but committing errors and learning to check one's results are part of the learning process. If the fear of errors is to be a deterrent, a child would never learn to walk.

It is the contention of this paper that understanding is achieved intuitively and that the logical presentation is at best a subordinate and supplementary aid to learning and at worst a decided obstacle. Intuition should fly the student to the conclusion, make a landing, and then perhaps call upon plodding logic to show the overland route to the same goals. If this contention is correct then the intuitive approach should be the primary one in introducing new subject matter *at all levels*. This recommendation may appear to be treason to mathematics, but let us withhold judgment.

3. The Historical Evidence. Though no air-tight case for the intuitive approach can be made from a study of the historical growth of mathematics a brief survey seems to offer some compelling arguments.

The first deductive structure was Euclid's *Elements*. Euclidean geometry, however, did not come into being in this form. It took 300 years, the period from Thales to Euclid, of exploration, fumbling, vague and even incorrect arguments before the *Elements* could be organized. Even this structure, intended to be strictly logical, rests heavily on intuitive arguments, pointless and even meaningless definitions, and inadequate proofs. That the logical structure can be devised after a subject is created and understood is not in question. What is relevant is that this deductive system came after the understanding was achieved. Moreover, it is no accident that Euclidean geometry was the first subject to receive any extensive mathematical development; the reason is that the intuition is readily applied to infer geometrical facts and the very figures suggest methods of proof.

A striking contrast is provided by the development of arithmetic and algebra.

Whole numbers and fractions and the operations with them were well accepted by the Egyptians and Babylonians, on an empirical basis, at least as far back as 2000 B.C. But irrational numbers, once their true character was recognized by the Pythagoreans, were not accepted by the classical Greeks as numbers. Why not? Because whole numbers and fractions had an obvious physical meaning whereas irrationals did not. The only intuitive meaning that one could attach to irrationals was that they represented certain geometrical lengths. What, then, did the Greeks do? They rejected irrationals as numbers and thought of them as lengths. In fact they converted all of algebra into geometry in order to work with lengths, areas and volumes that might otherwise have to be represented numerically by irrationals, and they even solved quadratic equations geometrically.

The progress that was made in the use of irrational numbers is due to the Alexandrian Greek civilization, which was a composite of the classical Greek, Egyptian and Babylonian civilizations, and to the Hindus and Arabs who were entirely empirically oriented. It was the Hindus who decided that $\sqrt{2}\sqrt{3}=\sqrt{6}$, and their argument was that these irrationals could be "reckoned with like integers," that is, like $\sqrt{4}\sqrt{9}=\sqrt{36}$. Irrational numbers were gradually accepted because of their utility and because familiarity breeds uncriticalness. The logical presentation of irrational numbers was not created until the 1870's.

Negative numbers, introduced by the practical-minded Hindus about 600 A.D., did not gain acceptance for 1000 years. The reason: they lacked intuitive support. The history of complex numbers is somewhat similar, though these did not appear until about 1540, and only about 200 years were required for these to be used somewhat freely. A remark of Gauss is very pertinent. As is well known, he was one of the men who discovered the geometrical representation of complex numbers, and about this he said in 1831 [1], "Here (in this representation) the demonstration of an intuitive meaning of $\sqrt{-1}$ is completely grounded and more is not needed in order to admit these quantities into the domain of the objects of arithmetic." Neither Descartes, Fermat, Newton, Leibniz, Euler, Lagrange, Gauss, or Cauchy could have given a definition of negative or complex numbers, or irrationals for that matter. Yet all of them managed to work with these numbers quite satisfactorily, to put it mildly, at least insofar as their times employed these numbers. In 1837 Hamilton did give the ordered couple definition of complex numbers in terms of real numbers but the logical development of the real number system itself was not constructed until the last part of the nineteenth century. The history of the entire complex number system is pertinent not only in itself but because algebra and analysis obviously utilize the number system and whatever basis there was for the latter had to serve as the basis for algebra and analysis.

The manner in which mathematics develops and is understood is beautifully exemplified by the history of the calculus. For the sake of brevity let us ignore the predecessors of Newton and Leibniz. The basic concept of the calculus is, of course, the instantaneous rate of change of a function, that is, the limit of

$\Delta y/\Delta x$ as Δx approaches 0. Where it was physically appropriate Isaac Newton thought of the limit in question as a velocity or as an acceleration, and he made great use of this fact in solving physical problems. But Newton experienced insuperable difficulties in explaining how he obtained the derivative from $\Delta y/\Delta x$. Because Δy approaches 0 when Δx does, he had to account for the fact that the quotient approached a definite number. Newton wrote three papers on the calculus and put out three editions of his famous *Mathematical Principles of Natural Philosophy*, and in each of these publications he made different explanations. In his first paper he says that his method is "shortly explained rather than accurately demonstrated." In his second paper he changed some terminology so as "to remove the harshness from the doctrine of indivisibles," but the logic is no more perspicuous. In the third paper Newton says "in mathematics minutest errors are not to be neglected." And then he gives a definition of the derivative, or fluxion as he called it, which supposedly shows that a fluxion is a precise concept. "Fluxions are, as near as we please, as the increments of fluents generated in times, as equal and as small as possible, and to speak accurately, they are in the prime ratio of nascent increments; yet they can be expressed by any lines whatever, which are proportional to them."

In the first and third editions of the *Principles* Newton says, "Ultimate ratios in which quantities vanish, are not, strictly speaking, ratios of ultimate quantities, but limits to which the ratios of these quantities decreasing without limit, approach, and which, though they can come nearer than any given difference whatever, they can neither pass over nor attain before the quantities have diminished indefinitely." He says further, "by the ultimate ratio of evanescent quantities is to be understood the ratio of the quantities, not before they vanish, nor after, but that with which they vanish." There are other statements by Newton in the published versions of his works which differ from the above. Clearly Newton struggled hard to define the derivative but scarcely succeeded in formulating a precise concept.

Leibniz worked not with the ratio $\Delta y/\Delta x$ and its limit but with differentials dx and dy which, he said, though not zero were not ordinary numbers. They were geometrically the differences in abscissa and ordinate, respectively, of two "infinitely near points." He too published many papers in which he tried to explain the meaning of the ratio dy/dx. Concerning his first paper on the calculus, published in 1684, even his friends, the Bernoulli brothers, said it was "an enigma rather than an explication."

Other papers and efforts to clarify his ideas did not accomplish any more. In a letter to Wallis Leibniz says: "It is useful to consider quantities infinitely small such that when their ratio is sought, they may not be considered zero but which are rejected as often as they occur with quantities incomparably greater. Thus if we have $x+dx$, dx is rejected. But it is different if we seek the difference between $x+dx$ and x. Similarly we cannot have xdx and $dxdx$ standing together. Hence, if we are to differentiate xy we write $(x+dx)(y+dy)-xy=xdy+ydx+dxdy$. But here $dxdy$ is to be rejected as incomparably less than $xdy+ydx$.

Thus in any particular case, the error is less than any finite quantity."

In the absence of satisfactory definitions he resorted to analogies to explain his differentials. At one time he referred to dy and dx as momentary increments or as vanishing or incipient magnitudes. These are Newtonian phrases. By way of additional explanation he said that as a point adds nothing to a line so differentials of higher order, e.g., $dxdx$, add nothing to dx. Alternatively dx is to x as a point to the Earth or as the radius of the Earth to that of the heavens. There are many other statements by Leibniz which are equally obscure.

There were many attacks on Leibniz's and Newton's work. Newton did not respond but Leibniz did. He objected to "overprecise critics" and argued that we should not be led by excessive scrupulousness to reject the fruits of invention. The phrases infinitely large and infinitely small signify no more than quantities which one can take as great or as small as one wishes. And then he adds that one can use these ultimate quantities, the actual infinite and the infinitely small, as a tool much as the algebraists use the imaginary with great profit. He also said that if one prefers to reject infinitely small quantities, it was possible instead to assume them to be as small as one judges necessary in order that they should be incomparable and that the error produced should be of no consequence or less than any given magnitude.

Of course the successors of Newton and Leibniz were aware of the lack of rigor in the calculus. Euler in the first classic text on the calculus, his *Introductio in Analysin Infinitorum* (1748), and again in his *Institutiones Calculi Differentialis* (1755) and *Institutiones Calculi Integralis* (1768–70), Lagrange in his *Théorie des fonctions analytiques* (1797) and d'Alembert in his article *Limite* in the *Encyclopédie* all struggled manfully but futilely to clarify the basic concepts of the calculus. Even Cauchy, the founder of rigor, gave definitions in his *Cours d'Analyse Algébrique* (1821) that would be considered loose and intuitive today. For example he says, "When the successive values attributed to a variable approach indefinitely a fixed value so as to end by differing from it by as little as one wishes, this last is called the limit of all the others." He defines continuity essentially by the requirement that the numerical value of the difference $f(x_0+\alpha)-f(x_0)$ decrease indefinitely with that of α. Because he was far more intuitive than rigorous Cauchy failed to distinguish between continuity and differentiability and even after his attention was called to this fact he persisted for twenty years in using differentiability where he had assumed only continuity. Clearly Cauchy's own rigor was beyond his comprehension. Cauchy also failed to recognize the necessity for the uniform convergence of series in order to integrate series term-by-term and to assert that the sum of a convergent series of continuous functions is continuous. Nor did he question that a function of two independent variables has a limit in both variables if it has a limit in each variable separately.

It is interesting that with respect to continuity and differentiability most texts of the nineteenth century including those written by the best mathematicians either followed Cauchy or "proved" that continuity implied differentiabil-

ity. This is why the mathematical world was shocked when Weierstrass in 1872 produced the example of a function that is continuous for all real values of x but has no derivative at any value of x. Luckily this example came late in the development of the calculus for, as Emile Picard said in 1905 [8], "If Newton and Leibniz had known that continuous functions need not necessarily have a derivative, the differential calculus would never have been created."

In view of the vague, unclear, and even incorrect foundations of the calculus one might expect that the subject would collapse. But before Weierstrassian rigor became known through his lectures at Berlin in the 1870's not only had the calculus been extended and applied but the subjects of ordinary and partial differential equations, the calculus of variations, differential geometry, and the theory of functions of a complex variable had been erected on the calculus. How did the mathematicians achieve these tremendous victories? Clearly they thought intuitively.

We could examine the developments in projective geometry, non-Euclidean geometry and other areas but the story would be about the same. One can safely say that no proof given up to at least 1850 in any area of mathematics, except in the theory of numbers, and even there the logical foundation was missing, would be regarded as satisfactory by the standards of 1900, to say nothing about today's standards. Yet the mathematics created by the men was surely understood by them. The history teaches us then that the intuition of great men is far more successful than their logic.

One could of course argue that the growth of mathematics may indeed have proceeded as described but now that we have the proper logical structures for the number system, algebra, analysis and the various branches of geometry we need not ask students to repeat the fumblings of the masters. We can give them the correct approaches and they will understand them. This argument can be countered with the fact that many mathematicians did try to build logical foundations for the various subjects—witness Euler, Lagrange and Cauchy in the calculus—and their failure to do so ought to be some evidence that the logical approaches are not easy to grasp. Of course our students are superior to the best mathematicians of the past.

There is not much doubt that the difficulties the great mathematicians encountered are precisely the stumbling blocks that students experience and that no attempt to smother these difficulties with logical verbiage will succeed. If it took mathematicians 1000 years from the time that first class mathematics appeared to arrive at the concept of negative numbers, and it did, and if it took another 1000 years for mathematicians to accept negative numbers, as it did, we may be sure that students will have difficulties with negative numbers. Moreover, the students will have to master these difficulties in about the same way that the mathematicians did, by gradually accustoming themselves to the new concepts, by working with them and by taking advantage of all the intuitive support that the teacher can muster.

These conclusions have been reached by many great mathematicians who

have concerned themselves with pedagogy. Poincaré said [12], "The zoologists maintain that in a brief period the development of the embryo of an animal recapitulates the history of its ancestors of all geological epochs. It appears that it is the same in the development of the mind. The task of the educator is to make the mind of the child go through what his fathers have experienced, to pass rapidly through certain stages but not to omit any. For this purpose, the history of the science ought to be our guide." Even Hilbert, the founder of modern axiomatics, granted the high pedagogic and heuristic value of the genetic method [2].

But we shall not insist on the evidence of history. There are other weighty arguments.

4. The Distortions of the Deductive Approach to Mathematics. Far from being the pedagogically sound representation of mathematics, the deductive approach introduces distorted views of the subject. First of all, mathematics is primarily a creative activity, and this calls for imagination, geometric intuition, experimentation, judicious guessing, trial and error, the use of analogies of the vaguest sort, blundering and fumbling. Even when a mathematician is convinced that a result must be correct he must still create to find the proof. As Gauss put it, "I have got my result but I do not know yet how to get it." Every mathematician knows that the hard work, tribulations and real thinking are required by, and the sense of achievement derives from, the creative effort. Writing up the final deductive formulation is a boring task.

Creativity presupposes flexibility in solving problems and any ideas from any domain of mathematics should be entertained whether or not they fall within the confines of a particular axiomatic structure. The latter, in fact, acts as a straitjacket on the mind.

What does logic contribute to the creation of concepts? Suppose one wishes to define the curvature of a surface. This definition is not arrived at by a deduction from axioms. It requires some deep insight to appreciate that this concept can be effectively represented by the product of the maximum and minimum curvatures of the curves through a point on the surface. As a matter of fact this definition was created by Euler who paid no attention to axiomatics.

Some of the greatest ideas in mathematics are not at all a matter of logic. Perhaps the best example is the realization that non-Euclidean geometry is applicable to physical space. The logical side, namely, pursuing the consequences of assuming a non-Euclidean parallel axiom, was a relatively simple task and was performed by Saccheri, Lambert, Legendre, Schweikart, Taurinus and many others. But it was Gauss who first recognized that these new geometries are as applicable as Euclidean geometry. The consequences for mathematics were as revolutionary as the very creation of mathematics itself.

It is true that some mathematicians, for example, Weierstrass in part of his work, Peano and Frege, produced rigorous theorems or axiomatic deductive structures. But in this work they were only reformulating what was already known and their goal was to rigorize what was well understood. New ideas were

never obtained in this manner. Logic discovers nothing, neither the statement of a theorem nor its proof, even in the construction of axiomatic formulations of known results. Thus the concentration on the deductive approach omits the real activity. The logical formulation does dress up this activity but conceals the flesh and blood. It is like the clothes which make the woman but are not the woman. It is the last act in the development of a branch of mathematics and, as one wise professor put it, when this is performed the subject is ready for burial. Logic may be a standard and an obligation of mathematics but it is not the essence.

The student should be creating mathematics. Of course he will be re-creating it and with the aid of a teacher. This recreativity on the part of the student is more popularly termed discovery today. Every teacher professes to espouse discovery. The student can be gotten to do this if he is allowed to think intuitively but he cannot be expected to discover within the framework of a logical development that is almost always a highly sophisticated and artificial reconstruction of the original creative work.

The logical version is a distortion of mathematics for another reason. The concepts, theorems and proofs emerged from the real world. It is the uses to which the mathematics is put that tell us what is correct. Thus we add fractions by finding a common denominator and not by adding numerators and adding denominators though we do multiply fractions by multiplying numerators and multiplying denominators. Likewise, the uses to which matrices are put determine that multiplication is to be noncommutative though we can devise purely mathematical multiplications of matrices that are commutative. After we have determined what properties mathematical concepts and operations must possess on the basis of the uses of these concepts and operations we *then* invent a logical structure, however artificial it must be, which yields these properties. Hence, the logic does *not* dictate the content of mathematics. The uses determine the logical structure. The logical organization is an afterthought. As Jacques Hadamard remarked, logic merely sanctions the conquests of the intuition. Or, as Weyl put, "logic is the hygiene which the mathematician practices to keep his ideas healthy and strong."

In fact, if a student is really bright and he is told to cite the commutative law to justify, say $3\cdot 4=4\cdot 3$, he may very well ask, Why is the commutative law correct? The true answer is, of course, that we accept the commutative law because our experience with groups of objects tells us that $3\cdot 4=4\cdot 3$. In other words the commutative law is correct because $4\cdot 3=3\cdot 4$ and not the other way around. The normal student will parrot the words commutative law, and he will, as Pascal put it in his *Provincial Letters*, "fix this term in his memory because it means nothing to his intelligence."

The deductive development of a branch of mathematics is often so artificial that it is meaningless. No example is more pertinent than the deductive development of the real number system. There were good reasons to axiomatize the number system, but the introduction of fractions and negative numbers as

couples with special definitions of the operations with these couples and the introduction of the irrationals by Cantor sequences or Dedekind cuts, clever as they may be, are so artificial, trumped-up and foreign to the intuitive meaning and uses of these numbers as to preclude understanding.

In such developments -2 is often introduced as the number which when added to 2 gives 0 or, as the modern mathematics texts put it, -2 is the unique additive inverse to 2. Such a definition induces no more understanding of -2 than the statement, anti-matter is that which added to matter produces a vacuum, gives any understanding of anti-matter. One doesn't learn even about dogs from a definition of dogs.

Poincaré makes this point too [9]. "In becoming rigorous mathematical science assumes a character so artificial as to strike every one. It forgets its historical origins; we see how the questions can be answered, but we no longer see how and why they were put."

Poincaré also notes [12] that in building up the number system from the integers there are many different constructions one can make. Why do we take one rather than another? "The choice is guided by the recollection of the intuitive notion in which this construction took place; without this recollection, the choice appears unjustified. But to understand a theory it is not sufficient to show that the path that one follows does not present obstacles; it is necessary to take account of the reasons that one chooses that path. Can one ever understand a theory if one builds it up right from the start in the definitive form that rigorous logic imposes, without some indications of the attempts which led to it? No; one does not really understand it; one cannot even retain it or one retains it only by learning it by heart."

Many teachers might retort that the student has already learned the intuitive facts about the number system and is now ready for the appreciation of the deductive version, which exemplifies mathematics. If the student really understands the number system intuitively the logical development will not only not enhance his understanding but will destroy it. As an example of mathematical structure no poorer choice could be made because the construction is so contrived. The development is so full of details and so stilted that it not only stultifies the mind but obscures the real ideas. Yet just this topic has now become the chief one in high school and college mathematics courses.

Actually this deductive approach is even misleading. In extending the number system from the natural numbers to the various other types we insist that the commutative and associative properties of the operations be retained. Why do we insist on these properties? We teachers know that the uses of the numbers call for these properties but the student gets the impression that these are necessary properties of all mathematical quantities. Why then do we not extend the order properties to complex numbers and the commutative property to matrices? The logical approach gives the student an entirely false impression of how mathematics develops.

The insistence on a deductive approach deceives the student in another

way. He is led to believe that mathematics is created by geniuses who start with axioms and reason directly from the axioms to the theorems. The student feels humbled and baffled, but the obliging teacher is fully prepared to demonstrate genius in action. Perhaps most of us do not need to be told how mathematics is created but it may help to listen to the words of Felix Klein [3]. "You can often hear from non-mathematicians, especially from philosophers, that mathematics consists exclusively in drawing conclusions from clearly stated premises; and that in this process, it makes no difference what these premises signify, whether they are true or false, provided only that they do not contradict one another. But a person who has done productive mathematical work will talk quite differently. In fact those people are thinking only of the crystallized form into which finished mathematical theories are finally cast. The investigator himself, however, in mathematics as in every other science, does not work in this rigorous deductive fashion. On the contrary, he makes essential use of his imagination and proceeds inductively aided by heuristic expedients. One can give numerous examples of mathematicians who have discovered theorems of the greatest importance which they were unable to prove. Should one then refuse to recognize this as a great accomplishment and in deference to the above definition insist that this is not mathematics? After all it is an arbitrary thing how the word is to be used, but no judgment of value can deny that the inductive work of the person who first announces the theorem is at least as valuable as the deductive work of the one who first proves it. For both are equally necessary, and the discovery is the presupposition of the later conclusion."

The deductive approach produces practical complications. If a student has to show, for example, that $4ab(ab+3ac)=4a^2b^2+12a^2bc$ and if he has to justify each step, he will have to think carefully and give reasons for so many steps that he will take minutes to do what he should do almost automatically on the basis of experience with numbers. It is far preferable that the student should become so familiar with the basic properties such as distributivity, commutativity and associativity that he does not realize he is using them. Likewise many students of calculus have learned (by heart) the proof that a continuous function on a closed interval has a maximum and a minimum but cannot find the maxima and the minima of simple functions.

We should be grateful that students accept unquestioningly facts that seem entirely reasonable to them whether on the basis of experience with numbers or intuitive arguments. In fact we should do all we can to make the elementary operations so habitual that students do not have to think about them any more than one thinks when he ties his shoelaces. If students do not see readily that $3\cdot x=x\cdot 3$, it is not because they lack familiarity with the commutative principle but rather because they fail to understand that x is just a number. (Of course, one should say, x is a placeholder for a number.) When the time to teach a noncommutative operation arrives, then commutativity can be stressed.

The need to make some of the work automatic was stressed by a man who certainly understood the role of axiomatics. Alfred North Whitehead says

[15], "It is a profoundly erroneous truism, repeated by all copybooks, and by eminent people when they are making speeches, that we should cultivate the habit of thinking of what we are doing. The precise opposite is the case. Civilization advances by extending the number of important operations which we can perform without thinking about them. Operations of thought are like cavalry charges in a battle—they are strictly limited in number, they require fresh horses and must only be made at decisive moments."

Modern texts are not content just to present mathematics deductively. They incorporate the rigor that meets the professional's standards. Thus the deductive approach to geometry now requires that the axioms overlooked by Euclid, the order axioms for example, be included to justify the steps. The consequence is that a host of trivial theorems must be proved before one reaches the significant ones. Thus the student must prove that there is a unique midpoint for each line segment and that there is an inside and an outside of a triangle. Even worse is the fact that many of the theorems are more obvious than the axioms used to establish them. Hence, the less obvious is used to prove the more obvious. But as far as the student is concerned the whole point of proof is just the reverse. Students will question what is being accomplished and perhaps even wonder whether we teachers are sane. That for two thousand years Euclidean geometry, as formulated by the presumably careless or naive Euclid, was regarded by the best mathematicians as the paradigm of rigor, bears no weight with the advocates of precise axiomatics. Today's students, we are apparently supposed to believe, are sharper and will not be satisfied with proofs that fail to mention details whose absence no one noticed for so many centuries.

Poincaré struck at this very folly [12]. "When a student commences seriously to study mathematics, he believes he knows what a fraction is, what continuity is, and what the area of a curved surface is; he considers as evident, for example, that a continuous function cannot change its sign without vanishing. If, without any preparation, you say to him: No, that is not at all evident; I must demonstrate it to you; and if the demonstration rests on premises which do not appear to him more evident than the conclusion, what would this unfortunate student think? He will think that the science of mathematics is only an arbitrary accumulation of useless subtleties; either he will be disgusted with it or he will amuse himself with it as a game and arrive at a state of mind analogous to that of the Greek sophists."

The rigorous approach requires such a multitude of minor theorems that the larger features of the subject fail to stand out. As Poincaré put it [10], "In the edifices built up by our masters, of what use is it to admire the work of the mason if we cannot comprehend the plan of the architect? Now pure logic cannot give us the appreciation of the total effect; this we must ask of the intuition."

In many areas the present emphasis on the logical approach is sheer hypocrisy. What mathematician uses the logical development of the complex number system to justify his operations with real or complex numbers? Yet this is what

is taught to students as the way to learn the "truth" about numbers. How many mathematicians have ever satisfied themselves that $\sqrt{2}^{\sqrt{3}}$ is defined in the theory of irrational numbers or even that $\sqrt{2}\sqrt{3}=\sqrt{3}\sqrt{2}$? How many have ever worked through a rigorous development of Euclidean geometry (as opposed to the pseudo-rigorous developments found in modern texts)? Felix Klein did not hestitate to admit [5], "To follow a geometrical argument purely logically without having the figure on which the argument bears constantly before me is for me impossible."

As a matter of fact the attempt to be completely deductive ensnares the teacher in a trap. It is often necessary to include a proof which even the rigor-oriented teachers concede to be too difficult for the student, such as the proof of the formula for the area of a circle in plane geometry. Many texts evade the issue by adopting an axiom. As a consequence numerous elementary geometry texts contain as many as 70 or 80 axioms. Surely if one can adopt axioms at will there is no need to prove anything. The only lesson the student will learn from such presentations is that if he is stuck he can adopt an axiom. The mathematics teacher can no more afford to be profligate with axioms than to be parsimonious. Likewise in the presentation of the real number system the high school texts proceed axiomatically from the natural numbers. But when they get to the irrational numbers, whose logical development the authors recognize to be too difficult for the student, they resort to the number line and speak of points which have no numbers assigned to them. These are designated by the irrational numbers. If the logical presentation of the rational numbers had any value it is dissipated by this meaningless introduction of the irrationals.

One of the gravest defects in the teaching of mathematics is the lack of motivation. Mathematics proper, as Weyl described it, has the inhuman quality of starlight, brilliant and sharp, but cold. Consequently very few students are attracted to the subject. In fact most of those taking high school and college mathematics do so because it is required or because they are prospective scientists or engineers. These students would prefer to learn more about the fruits than the roots of mathematics. Proper pedagogy requires that these students be shown why they should be studying particular topics and subjects. To assure them that the material will prove useful at some later time is hardly an incentive to take it seriously. The mathematician or the rare student who finds intellectual challenge or aesthetic satisfactions in the subject may be intrigued to learn that there are only five regular polyhedra. But very few students are excited by this fact. As far as they are concerned the world would be just as well off if there were an infinite number of them. As a matter of fact there is an infinite number of regular polygons and no one seems depressed by this fact.

Although the subject of motivation is a vast one its relevance here is simply that it is relatively easy to give a genuine or significant motivation to a mathematical topic when this is introduced intuitively or heuristically because historically there were significant motivations, whereas it is very difficult to do so

in a logical presentation because the latter is many stages removed from reality and, as we have already pointed out, is often artificial. How does one motivate the concept of a fraction when it is to be introduced as an ordered couple of natural numbers? How can a student see the point of the $\epsilon-\delta$ definition of continuity if this is his introduction to the concept? In fact the logical approach often destroys the motivation. One may motivate the integral as the method of finding the area under a curve. But if one defines the area as an integral one begs the whole question of doing something significant with the integral.

Many texts and teachers claim that they do provide motivation even for a logical approach. Thus they "motivate" the introduction of negative, irrational and complex numbers by stating that we wish to solve equations such as $x^2+2=0$. But for students who have no reason to solve even $x-2=0$, the challenge of solving $x^2+2=0$ certainly isn't exciting. Moreover, the bright student can come back at the teacher and ask, "Why can't we solve $5/x=0$ by introducing ∞ as a number?" If we can invent definitions to operate with $\sqrt{-2}$, we can invent definitions to operate with ∞.

The presentation of theorems without the motivation robs the student of insight. Even on the somewhat advanced level where we deal with students who have some leaning toward mathematics, to present theorems without the motivation, whether it does or does not kill off the interest in mathematics, certainly leaves them with no more than a meaningless collection of theorems and proofs and without the power to think for themselves. Thus in linear algebra texts the subject of eigenvalues of matrices is always treated. I have not found one text which indicates why one wants to learn anything about the eigenvalues. The simultaneous reduction of two quadratic forms to sums of squares is another seemingly meaningless topic. At least the origin in mechanical problems might be suggested. Equivalent, congruent, and similar matrices are treated for no apparent reason. (See the article by R. J. Jarvis: A Case for Applications of Linear Algebra and Group Theory, this MONTHLY, 73, 1966, 654–656.)

5. The Role of Logic in Pedagogy. In view of the many pedagogical shortcomings in the logical approach to mathematics it is not surprising that many perceptive mathematicians (there are nonperceptive ones) have spoken out against the logical approach. Descartes deprecated logic in rather severe language. "I found that, as for Logic, its syllogisms and the majority of its other precepts are useful rather in the communication of what we already know or . . . in speaking without judgment about things of which one is ignorant." Roger Bacon said, "Argument concludes a question but it does not make us feel certain, or acquiesce in the contemplation of a truth, except the truth also be found to be so by experience." Pascal pointed out that "Reason is the slow and tortuous method by which those who do not understand the truth discover it."

Is there then no role for logic or proof? Should it be rejected all together? Not at all. The first approach to any subject should indeed be intuitive. As

Poincaré put it [9], "I have already had occasion to insist on the place intuition should hold in the teaching of the mathematical sciences. Without it young minds could not make a beginning in the understanding of mathematics; they could not learn to love it and would see in it only a vain logomachy; above all without intuition they would never become capable of applying mathematics. . . .

"We need a faculty which makes us see the end from afar and intuition is this faculty."

Proof should enter but only gradually. Moreover, the level of rigor must be suited to the level of the student's mathematical development. The proof need only convince the student. The capacity to appreciate rigor is a function of the mathematical age of the student and not of the age of mathematics. This appreciation is acquired gradually, and the student must have the same freedom to make intuitive leaps that the mathematicians had. Rigor will not refine an intuition that has not been allowed to function freely. Proofs of whatever nature should be invoked only where the students think they are required. A proof is meaningful when it answers doubts. Felix Klein has stressed this point [6]: "It is my opinion that in teaching it is not only admissible, but absolutely necessary, to be less abstract at the start, to have constant regard for applications, and to refer to the refinements only gradually as the student becomes able to understand them. This is, of course, nothing but a universal pedagogical principle to be observed in all mathematical instruction." As Professor Max M. Schiffer of Stanford University has stated it, "Never put logical carts before heuristic horses."

The level of rigor can, of course, be advanced as the student progresses. Poincaré makes this point too [12]. "On the other hand, when he is more advanced, when he becomes familiar with mathematical reasoning and his mind will be matured by this very experience, the doubts will be born of themselves and then your demonstration will be well received. It will awaken new doubts and the questions will arise successively to the child as they arose successively to our fathers to the point where only perfect rigor can satisfy him. It is not sufficient to doubt everything; it is necessary to know why one doubts."

6. Why is the Deductive Approach Favored? Despite the pedagogical defects of the deductive approach, the criticisms of big mathematicians, and the claims of many mathematicians that they do teach discovery, the prevailing practice, if I may judge from the textbooks and hundreds of talks with professors, is to present mathematics rigorously and to emphasize the axiomatic method. Indeed this is the essence of the so-called reform known as "modern mathematics" or the "new mathematics." Why do teachers use this approach?

There is no doubt that some teachers actually believe that the axiomatic deductive presentation is the essence of mathematics. Whether they acquired this limited view through the instruction they themselves received or have been induced to adopt it because the textbooks now favor it, they are at least sincere if not effective pedagogues. One has the sneaking suspicion that a few

teachers enjoy presenting the familiar number system in the recondite axiomatic form because they understand the simple mathematics it represents and yet can appear to be presenting profound mathematics. Certainly much of the rigor in modern texts comes from limited men who seek to conceal shallowness by giving a facade of profundity to the obvious and from pedants who mask their pedantry under the guise of rigor.

Many young teachers believe that now that we have the correct, polished version of mathematics it is sufficient to give the axiomatic or rigorous approach and that students will absorb it. These very same teachers would have been swamped by such a presentation but having learned the correct version they can no longer recall and appreciate the difficulties they encountered in learning the rigorous versions.

Some teachers, knowing the rigorous proofs, feel uneasy about presenting a convincing argument which they, at least, know is incomplete. But it is not the teacher who is to be satisfied; it is the student. Good pedagogy demands such compromises.

Other teachers want to give students the whole truth at once so that they should not have to unlearn what they once learned. But one cannot teach even English or History by starting at the top. The A that a high school student might earn for an English composition would most likely be rated C at the college level.

For whatever reason teachers insist on presenting to young people a modern rigorous proof they are deceiving themselves. There is no ultimate rigorous proof. This fact derives from the very way in which mathematics develops. Felix Klein has described it [4]: "In fact, mathematics has grown like a tree, which does not start at its tiniest rootlets and grow merely upward, but rather sends its roots deeper and deeper at the same time and rate that its branches and leaves are spreading upward. . . . *We see, then, that as regards the fundamental investigations in mathematics, there is no final ending, and therefore on the other hand, no first beginning, which could offer an absolute basis for instruction.*" Poincaré expressed a similar view. There are no solved problems; there are only problems that are more or less solved. Mathematics is as correct as human beings are and humans are fallible.

At no time in the history of mathematics have we been less certain of what rigor is. Hence no proof is really complete, and the teacher must compromise in any case. It would be interesting to know how many teachers are aware that set theory, which they now regard as the indispensable beginning to any rigorous approach to mathematics, has been the source of our deepest and thus far insuperable logical difficulties [16]. Those who are not aware of the foundational problems might at least note the words of Hermann Weyl [14]: "The question of the ultimate foundations and the ultimate meaning of mathematics remains open; we do not know in what direction it will find its final solution nor even whether a final objective answer can be expected at all. 'Mathematizing' may well be a creative activity of man, like language or music, of primary

originality, whose historical decisions defy complete objective rationalization."

Many teachers favor a logical presentation, particularly, an abstract one, such as group theory, because it is supposed to be efficient. They are under the impression that if a student is taught abstract groups he will in one swoop learn the properties of the rational, real and complex numbers, matrices, congruences, transformations and other topics. But of course a student who learns group theory could not on this basis add fractions. Nor does offering an example or two of a group save the day. The concrete cases must be thoroughly understood *before* one introduces an abstract development which unifies several concrete ones. To introduce as examples concrete material which is as yet unfamiliar to the student is of no help at all in making the abstract notion clearer. In every case learning proceeds from the concrete to the abstract and not vice versa.

However, the major reason for the popularity of the axiomatic rigorous approach is that it is easier to teach. The entire body of material is laid out in a clear, clean-cut sequence and all the teacher has to do is repeat it. He has but to offer a canned body of material. I have heard teachers complain that many students, particularly engineers, wish to be told how to perform the processes they are asked to learn and then want to hand back the processes. But the teachers who teach the logical presentation because it avoids such difficulties as teaching discovery, leading students to participate in a constructive process, explaining the reasons for proceeding one way rather than another, and finding convincing arguments, are more reprehensible than the students who wish to avoid thinking and prefer just to repeat mechanically learned processes. Postulating properties has the advantage, as Bertrand Russell put it, of theft over honest toil. Pedagogically it is worse because the theft produces no gain in understanding. The logical approach to teaching is reminiscent of a reply that Samuel Johnson gave to a man who asked Johnson for further explanation of some argument he had given. Johnson barked, "I have found you an argument but I am not obliged to find you an understanding."

Many of us know the story of the professor who was presenting a logical proof to his class, got stuck in the course of the proof, went over to the corner of the blackboard where he drew some pictures, erased the pictures, and then continued the proof. Whether the import of this story for pedagogy has been noted is doubtful.

Many mathematicians prefer to present rigorous axiomatic approaches which, for example, use a minimal set of axioms, because they favor their own professional interest at the expense of the student. Even if such systems can be made understandable to young people the time required to teach them could be spent on more significant material. In this matter as well as in presenting sophisticated rigorous proofs they are using the classroom to challenge themselves. These professors are serving themselves rather than the students not only in the form in which they present the various subjects but also in the premature teaching of abstractions such as abstract algebraic concepts, linear

vector spaces, finite geometries, set theory, symbolic logic and functional analysis, because these subjects lend themselves to axiomatic treatments. Is it any wonder that students become alienated and question the relevance of what they are being taught?

There are many indications that professors who present rigorous material are really uncertain as to the wisdom of doing so. A number of calculus books begin with rigorous definitions and theorems, for example, those concerning limits and continuity, and then never refer to this material. Thereafter they use the cookbook presentation. The most charitable view of such books is that the authors wish to ease their own consciences or to give the students some idea of what rigor means. Perhaps an unfairly severe view is that these books offer only a pretense of rigor in order to appeal to both markets, the one that demands rigor and the one that is satisfied to teach mechanical procedures.

Other texts adopt another "compromise." In the body of the text the presentation is mechanical with perhaps an occasional condescension to an intuitive explanation. The real "explanation" is given in rigorous proofs but these are put in appendices and presented so compactly that they are certain to be totally ununderstandable to the student. However, the authors have salved their consciences. Such books are no different from the old mechanical presentations. They do contribute to understanding in one respect, namely, they show that competent mathematicians are inept in pedagogy.

Perhaps, after all, there is some merit to the logical approach to mathematics. It has been said of rigor that "The virtue of a logical proof is not that it compels belief but that it suggests doubts and the proof tells us where to concentrate our doubts." Or as Bertrand Russell put it [13], "It is one of the chief merits of proofs that they instill a certain scepticism as to the result proved." Lebesgue pointed out another value of rigorous proof [7]. "Logic makes us reject certain arguments but it cannot make us believe any argument." One must respect but suspect mathematical proofs. Since one of the main objectives of mathematics education is to instill scepticism in the student, he is deriving at least one benefit from the current logical extravaganzas.

References

1. C. F. Gauss, Werke II, 177.

2. D. Hilbert, Grundlagen der Geometrie, 7th ed., Teubner, Leipzig, 1930, p. 242.

3. Felix Klein, Elementary Mathematics from an Advanced Standpoint, I, Macmillan, New York, 1932, and Dover, New York, 1945, p. 207.

4. ———, *Ibid.*, p. 15.

5. ———, Zur nicht-Euklidische Geometrie, Math. Ann., 37 (1890) 571, Ges. Math. Abh., I, 381.

6. ———, Gesammelte mathematische Abhandlungen, II, 231.

7. Henri Lebesgue, Leçons sur L'Intégration, 2nd ed., Gauthier-Villars, Paris, 1928, p. 328.

8. Émile Picard, Sur le Développement de L'Analyse et ses Rapports avec diverses Sciences, Gauthier-Villars, Paris, 1905, p. 5.

9. H. Poincaré, The Foundations of Science, The Science Press, Lancaster, 1913, p. 217.

10. ———, *Ibid.*, 435.

11. H. Poincaré, See also *ibid.*, 216.
12. ———, La Logique et L'Intuition, L'Enseignement Mathématique, 1 (1899) 157–162 = Oeuvres, XI, 129–134.
13. Bertrand Russell, The Principles of Mathematics, George Allen and Unwin, London, 1903, p. 360.
14. Hermann Weyl, Obituary David Hilbert, Obituary Notices of Fellows of the Royal Society, 4 (1944) 547–553, Ges. Abh., IV., p. 126.
15. A. N. Whitehead, An Introduction to Mathematics, Henry Holt, 1911, p. 61.
16. R. L. Wilder, The role of the axiomatic method, this MONTHLY, 74 (1967) 115–127.

MATHEMATICAL NOTES

EDITED BY DAVID DRASIN

Manuscripts for this Department should be sent to David Drasin, Division of Mathematical Sciences, Purdue University, Lafayette, IN 47907.

A COROLLARY TO THE GELFAND-MAZUR THEOREM

H. A. SEID, University of California, Irvine

We wish to establish the following result as a corollary to the Gelfand-Mazur Theorem:

Let $A \neq 0$ be a complex commutative normed algebra such that $\|xy\| = \|x\| \, \|y\|$ for all $x, y \in A$. Then A is isomorphic and isometric to the complex field.

Note that there is no assumption that A contains an identity. In fact the major important step in the construction of the proof is to demonstrate that under the given norm condition on A, the algebra has an identity. (If A were assumed to have an identity, then the above corollary would merely be a special case of a well-known result which appears in the literature. See for example [**3**, Cor. 1.7.3, p. 39]. Also under the assumption of an identity in A, Lorch has a proof of the above corollary [**1**, Th. 5-2, p. 129].)

We shall use an algebraic approach to allow us to demonstrate the existence of the identity in A. To obtain the result, the Gelfand-Mazur Theorem will be applied. The conclusion of the corollary follows directly.

Proof. By the norm condition, A has no zero divisors. Let $S = A - \{0\}$. Let $T = \{(a, s) \mid a \in A, s \in S\}$. As in Herstein [**2**, pp. 101–103] define an equivalence relation $\sim$ on T as follows: $(a, s) \sim (a_1, s_1)$ if and only if $s_1 a = s a_1$. Denote the equivalence class of (a, s) by a/s. We define $S^{-1}A = \{a/s \mid (a, s) \in T\}$. Then $S^{-1}A$ is the field of quotients of A under operations $+, \cdot$, now to be defined. Let $a/s, a_1/s_1 \in S^{-1}A$.

$$a/s + a_1/s_1 \equiv (s_1 a + s a_1)/s s_1 \tag{1}$$

$$(a/s) \cdot (a_1/s_1) \equiv a a_1/s s_1 = a_1 a/s_1 s = (a_1/s_1) \cdot (a/s). \tag{2}$$

Note that for every $s \in S$, s/s is the unit element for $S^{-1}A$. Let $\lambda \in \mathbf{C}$, $a/s \in S^{-1}A$. Then $\lambda(a/s) \equiv \lambda a/s$. Hence $S^{-1}A$ is a complex algebra. We norm $S^{-1}A$ as follows: for $a/s \in S^{-1}A$, $\|a/s\|' \equiv \|a\|/\|s\|$; $\|\cdot\|'$ is well defined by use of the norm condition on A and the definition of equivalence classes in $S^{-1}A$. We show that $\|\cdot\|'$ is a norm by demonstrating the triangle inequality. (The other conditions necessary are trivially true.) Let a/s, $a_1/s_1 \in S^{-1}A$.

$$\begin{aligned} \|a/s + a_1/s_1\|' &= \left\| \frac{s_1a + sa_1}{ss_1} \right\|' = \frac{\|s_1a + sa_1\|}{\|s_1s\|} \\ &\leqq \frac{\|s_1a\| + \|sa_1\|}{\|s_1s\|} = \frac{\|s_1\|\,\|a\| + \|s\|\,\|a_1\|}{\|s_1\|\,\|s\|} \\ &= \frac{\|a\|}{\|s\|} + \frac{\|a_1\|}{\|s_1\|} = \|a/s\|' + \|a_1/s_1\|'. \end{aligned}$$

Also $\|(a/s)\cdot(a_1/s_1)\|' = \|a/s\|'\|a_1/s_1\|'$. Let G be the completion of $S^{-1}A$. Then G is a field and a Banach algebra. By the Gelfand-Mazur Theorem, $G \approx \mathbf{C}$ and G is isometric to $\mathbf{C}$. Note that if $\phi\colon G \to \mathbf{C}$ is the isomorphism, $\phi(s/s) = 1 \in \mathbf{C}$ for $s \in S$. For $\lambda \in \mathbf{C}$, $\phi(\lambda s/s) = \lambda\phi(s/s) = \lambda \cdot 1 = \lambda$. Hence, $S^{-1}A \approx \mathbf{C}$.

Define $\beta\colon A \to S^{-1}A$ by $x \to xs/s$. It is easily seen that β is an injective algebra homomorphism. To see that β is surjective suppose $\lambda s/s \in S^{-1}A$ for $\lambda \in \mathbf{C}$. Let $x \in S \subset A$. $\beta(x) = xs/s = \lambda_1 s/s$ for some $\lambda_1 \in \mathbf{C}$ and $\lambda_1 \neq 0$. Then $(\lambda/\lambda_1)x \in A$ and $\beta((\lambda/\lambda_1)x) = \lambda s/s$. Hence under β, $A \approx S^{-1}A$. Thus A has an identity element ι. It follows easily that β is an isometry.

References

1. E. R. Lorch, Spectral Theory, Oxford University Press, New York, 1962.
2. I. N. Herstein, Topics in Algebra, Blaisdell, Waltham, Mass., 1964.
3. C. E. Rickart, General Theory of Banach Algebras, Van Nostrand, Princeton, N. J., 1960.

ON AITKEN'S Δ^2-METHOD

SIMEON REICH, The Technion, Haifa, Israel

In [1], p. 73, the following result is proved:

THEOREM 1. *Let $\{x_n\}$ be any sequence (of real numbers) converging to the limit s, such that the quantities $d_n = x_n - s$ satisfy $d_n \neq 0$ for n sufficiently large, and $d_{n+1}/d_n \to A$, where $|A| < 1$. Define $\{x_n'\}$ by*

$$x_n' = x_n - (\Delta x_n)^2/\Delta^2 x_n$$

(this is Aitken's method where $\Delta x_n = x_{n+1} - x_n$). Then $\{x_n'\}$ is actually defined for n sufficiently large and converges faster to s than the original sequence in the sense that

$$(x_n' - s)/(x_n - s) \to 0.$$

Although of no importance in the argument leading to this theorem, we remark that if the A given in the statement of Theorem 1 is not zero, the convergence of $\{x_n\}$ is said to be linear; otherwise it is said to be nonlinear. Now, since the result quoted is correct, it seems puzzling that in another book ([2], p. 348) the reader is told not to use Aitken's method in case the convergence is nonlinear. Perhaps we can resolve this "paradox" in the following manner: In Theorem 1, the rate of convergence of $\{x'_n\}$ is compared with that of $\{x_n\}$; but it is more reasonable to compare $\{x'_n\}$ with $\{x_{n+2}\}$, because x_{n+2} is already used in the definition of x'_n. If we do this, we obtain a slightly different result.

THEOREM 1'. *Under the hypotheses of Theorem* 1, *and with the additional assumption that* $A\neq 0$, $(x'_n-s)/(x_{n+2}-s)\to 0$.

Proof. We may write $d_{n+1}/d_n=A+\varepsilon_n$ for n sufficiently large, where $\varepsilon_n\to 0$, and then

$$(x'_n - s)/(x_{n+2} - s) = [\varepsilon'_n - 2\varepsilon_n(A-1) - \varepsilon_n^2] \\ /[((A-1)^2 + \varepsilon'_n)(A+\varepsilon_{n+1})(A+\varepsilon_n)]$$

converges to zero, where

$$\varepsilon'_n = A(\varepsilon_n + \varepsilon_{n+1}) - 2\varepsilon_n + \varepsilon_n\varepsilon_{n+1} \to 0.$$

What happens when $A=0$? Then

$$(x'_n - s)/(x_{n+2} - s) = [\varepsilon_n(\varepsilon_{n+1} - \varepsilon_n)]/[\varepsilon_n\varepsilon_{n+1}(1 - 2\varepsilon_n + \varepsilon_n\varepsilon_{n+1})].$$

Because $A=0$, $d_{n+1}/d_n=\varepsilon_n$ and therefore $\varepsilon_n\neq 0$ for n sufficiently large. This means that

$$(x'_n - s)/(x_{n+2} - s) = [1/(1 - 2\varepsilon_n + \varepsilon_n\varepsilon_{n+1})][1 - \varepsilon_n/\varepsilon_{n+1}].$$

Thus in this case,

$$(x'_n - s)/(x_{n+2} - s) \to 0$$

if $\varepsilon_n/\varepsilon_{n+1}\to 1$. That is if

$$(d_{n+1}/d_n)/(d_{n+2}/d_{n+1}) \to 1.$$

For example, this will indeed happen if we put $x_n=n^{-n}$, but will not happen if $x_n=2^{-n^2}$.

References

1. Peter Henrici, Elements of Numerical Analysis, Wiley, New York, 1964.
2. Anthony Ralston, A First Course in Numerical Analysis, McGraw-Hill, New York, 1965.

ON WEAKENING OF A POSTULATE OF H. WEYL

S. C. SAXENA, University of Akron

Given a vector space of dimension n, an affine space of dimension n may be defined by using axioms of H. Weyl. This is done as follows:

Let V be a vector space of dimension n defined over a field F. Let S be a nonempty set, and

$$\rho: S \times S \to V$$

a mapping having the following properties:

P. 1: For every three points P, Q, R in S,

$$\rho(P, Q) + \rho(Q, R) = \rho(P, R).$$

P. 2: For every $P \in S$, and $\alpha \in V$, there exists a unique point Q in S, such that $\rho(P, Q) = \alpha$.

The set S with the vector space V is called an affine space of dimension n [1].

The purpose of this note is to show that the postulate P. 2 can be replaced by the following weaker postulate:

P. 2′: There exists a point $P_0 \in S$ such that for each $\alpha \in V$, there exists a unique point $Q \in S$ such that $\rho(P_0, Q) = \alpha$.

Proof. In P. 1 first taking $P = Q = R$, we get $\rho(P, P) = 0$, then taking $P = R$, we get $\rho(P, Q) = -\rho(Q, P)$.

Given $P \in S$ and $\alpha \in V$, by P. 2′ there exists a unique $Q \in S$ such that $\rho(P_0, Q) = \alpha + \rho(P_0, P)$, i.e., such that $\rho(P, Q) = \rho(P, P_0) + \rho(P_0, Q) = \rho(P_0, Q) - \rho(P_0, P) = \alpha$. Thus P. 2 holds.

From a pedagogical point of view P. 2′ appears better than P. 2, since the point P_0 may be represented by zero vector of V and we can start coordinatization of S in terms of vectors of V, from there.

Reference

1. K. Nomizu, Fundamentals of Linear Algebra, McGraw-Hill, New York, 1966, pp. 259–260.

A RATIO-TYPE CONVERGENCE TEST

H. MILLER, DePaul University

In this note we introduce a ratio-type convergence test which properly contains the half of the Gauss Convergence Theorem dealing with convergence [1].

THEOREM 1. *If* $a_1 = 1$, $a_n > 0$ *for* $n \geqq 2$ *and* $a_{n+1}/a_n = 1 - d_n$, *then the convergence of*

$$\sum_{n=1}^{\infty} \left(1 - n^{-1} \sum_{k=1}^{n} d_k\right)^n$$

implies the convergence of $\sum_{n=1}^{\infty} a_n$.

Proof. We have $a_1=1$, $a_2=a_1(1-d_1)=(1-d_1)$, $a_3=a_2(1-d_2)=(1-d_1)\cdot(1-d_2)$, $\cdots$, and $a_n=(1-d_1)(1-d_2)\cdots(1-d_{n-1})$, $\cdots$. Therefore

$$\sum_{n=1}^{\infty} a_n = 1 + \sum_{n=1}^{\infty}\prod_{k=1}^{n}(1-d_k) \leqq 1 + \sum_{n=1}^{\infty}\left(n^{-1}\sum_{k=1}^{n}(1-d_k)\right)^n,$$

since the geometric mean of n positive numbers is less than or equal to the arithmetic mean. Therefore

$$\sum_{n=1}^{\infty} a_n \leqq 1 + \sum_{n=1}^{\infty}\left(1 - n^{-1}\sum_{k=1}^{n} d_k\right)^n,$$

completing the proof.

THEOREM 2. *If $a_1=1$, $a_n>0$ for $n\geqq 2$ and if $a_{n+1}/a_n=1-c/n+\theta_n/n^\lambda$ where $c>1$, if $\{\theta_n\}_{n=1}^{\infty}$ is bounded, and if $\lambda>1$, then*

$$\sum_{n=1}^{\infty}\left(1 - n^{-1}\sum_{k=1}^{n} d_k\right)^n$$

is convergent, where $d_k=1-a_{k+1}/a_k$. In other words, the test in Theorem 1 *contains the half of the Gauss convergence theorem dealing with convergence.*

Proof. For each k, $d_k=c/k-\theta_k/k^\lambda$. Hence

$$\begin{aligned}\left(1 - n^{-1}\sum_{k=1}^{n} d_k\right)^n &= \left(1 - cn^{-1}\sum_{k=1}^{n}(k^{-1} - \theta_k/ck^\lambda)\right)^n \\ &= (1 - cn^{-1}(\log n + c_n))^n,\end{aligned}$$

where $\{c_n\}$ is bounded. Now let δ be given satisfying $1<\delta<c$. There exists an N such that

$$\left(1 - n^{-1}\sum_{k=1}^{n} d_k\right)^n \leqq (1 - \delta n^{-1}\log n)^n$$

for all $n>N$. But $(1-\delta n^{-1}\log n)^n \leqq n^{-\delta}$ for all $n>N$, which follows from $\log(1-\delta n^{-1}\log n)\leqq -\delta n^{-1}\log n$. Therefore

$$\sum_{n=1}^{\infty}\left(1 - n^{-1}\sum_{k=1}^{n} d_k\right)^n$$

is convergent by the comparison test.

To complete the proof that the ratio-type convergence test described in Theorem 1 properly contains half of the Gauss convergence test, we submit the following example. Let $S=\{s_1, s_2, \cdots, s_j\}$ be a set consisting of j positive real numbers. One can construct sequences $\{a_n\}$ inductively in the following way: Let $a_1=1$ and let $a_{n+1}=sa_n$ for every positive integer n, where s is picked at random from the set S. We will show that if $\sum_{k=1}^{j} s_k<j$, then with probability one a sequence constructed as above is convergent.

The proof is as follows: for $\rho>0$

$$j^{-1}\sum_{k=1}^{j} s_k = 1-\rho$$

implies with probability one, by the Strong Law of Large Numbers [2], that there exists N such that $n^{-1}\sum_{k=1}^{n} d_k>\rho/2$ for all $n>N$. Therefore,

$$\sum_{n=1}^{\infty}\left(1-n^{-1}\sum_{k=1}^{n} d_k\right)^n$$

is convergent with probability one. Therefore, $\sum_{n=1}^{\infty} a_n$ is convergent with probability one.

If $j^{-1}\sum_{k=1}^{j} s_k<1$ and there exists a s_i in S such that $s_i>1$, then with probability one a sequence constructed as above is not "eventually monotonic." Therefore, the Gauss convergence test does not apply to almost all (probability one) such sequences. In fact, the whole scale of ratio tests will fail to show almost all of the above sequences are convergent [3].

References

1. K. Knopp, Theory and Application of Infinite Series, Blackie, London, 1948.
2. W. Feller, An Introduction to Probability Theory and Its Applications, Vol. II, Wiley, London, 1965, p. 233.
3. T. Bromwich, An Introduction to the Theory of Infinite Series, Macmillan, London, 1949, p. 42.

SOLUTIONS OF $\phi(x)=n$, WHERE ϕ IS EULER'S ϕ-FUNCTION

K. W. WEGNER AND S. R. SAVITZKY, Carleton College

There are simple formulas for evaluating Euler's ϕ-function. Finding values of x for $\phi(x)=n$ is more complicated. ($\phi(x)=1440$ has 72 solutions.)

R. D. Carmichael [Journal of Mathematics, 30(1908) 394–400] listed all the solutions of $\phi(x)=n$ for n up to 1000.

The computer has revealed the following errors in Carmichael's table:

For $n=768$, add solutions 1785 and 3570.

For $n=792$, delete 2384, add 2388.

For $n=888$, add 1043 and 2086.

For $n=960$, add 1309 and 2618.

For $n=972$, add 1467 and 2934.

All solutions have been calculated for n up to 1978. They may be obtained by writing to the Department of Mathematics, Carleton College, Northfield, Minnesota 55057.

RESEARCH PROBLEMS

EDITED BY VICTOR KLEE

In this Department the Monthly presents easily stated research problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Victor Klee, Department of Mathematics, University of Washington, Seattle, WA 98105.

WHICH ISOPERIMETRIC RATIOS ARE BOUNDED?

VICTOR KLEE, University of Washington

A subset of a Euclidean space is called a *polytope* provided that it is the convex hull of a finite set of points or, equivalently, is the bounded intersection of a finite number of closed halfspaces. A *d-polytope* is a polytope P that is d-dimensional. For $0 \leqq k < d$, a *k-face* of P is a k-dimensional set, necessarily a k-polytope, which is the intersection of P with a supporting hyperplane; the only *d-face* of P is P itself. For $0 \leqq k \leqq d$, let $\zeta_k(P)$ denote the sum of the k-measures of the various k-faces of P. Thus $\zeta_0(P)$ is the number of vertices of P, $\zeta_1(P)$ is the sum of the lengths of P's edges, $\zeta_2(P)$ is the sum of the areas of P's 2-faces, $\cdots$, $\zeta_{d-1}(P)$ is the $(d-1)$-dimensional surface area of P, and $\zeta_d(P)$ is the d-dimensional volume of P. (Note that if K is a k-face of P and F is the smallest flat containing K, then F may be regarded as a copy of Euclidean k-space E^k and the k-measure of K may therefore be taken in the sense of Lebesgue.)

Let I_d denote the set of all pairs (i, j) of distinct integers between 1 and d. For $(i, j) \in I_d$, the *(i, j)-isoperimetric ratio* of P is defined as

$$\rho_{i,j}(P) = \zeta_i(P)^{1/i}/\zeta_j(P)^{1/j}.$$

Note that $\zeta_k(\lambda P) = |\lambda|^k \zeta_k(P)$, so that $\rho_{i,j}(\lambda P) = \rho_{i,j}(P)$ for all $\lambda \neq 0$; hence the ratio $\rho_{i,j}(P)$ depends only on the shape and not on the size of P. The following is a precise statement of the title question:

For which triples (d, i, j) is the isoperimetric ratio $\rho_{i,j}(P)$ bounded above as P ranges over all d-polytopes?

For any such triple there is interest in determining the least upper bound

$$\beta(d, i, j) = \sup\{\rho_{i,j}(P) : P \text{ is a } d\text{-polytope}\}$$

and thus obtaining an analogue of the classical isoperimetric inequality (which asserts $\beta(d, d, d-1) = (d\omega_d^{1/d})^{-1/(d-1)}$, where ω_d is the volume of a d-dimensional spherical ball of unit radius).

Eggleston, Grünbaum, and Klee [2] show $\beta(d, i, j)$ is finite if $i = d$ or $i = d-1 > j$ or i is a multiple of j. Other results related to the finiteness of $\beta(d, i, j)$ appear in [2] and in Larman and Mani [3]. However, the finiteness of $\beta(d, i, j)$ is unsettled whenever $d-2 \geqq i > j \geqq 2$ and i is not a multiple of j. In particular, the finiteness of $\beta(d, d-2, d-3)$ is unsettled for all $d \geqq 5$. Though the supremum $\beta(d, d, d-1)$

is finite, it is not attained by any polytope; Grünbaum conjectures that all other finite bounds are attained.

Even when the number $\beta(d, i, j)$ is known to be finite, its exact value has been determined only for the cases corresponding to the classical isoperimetric inequality. In particular, the values of $\beta(d, d-1, 1)$ and $\beta(d, d, 1)$ are unknown for all $d \geqq 3$. Aberth [1] shows $\beta(3, 2, 1) \leqq (6\pi)^{-1/2}$. Melzak [4] conjectures $\beta(3, 3, 1) \leqq 2^{-2/3}3^{-11/6}$, with equality only for a right prism whose base is an equilateral triangle having side-length equal to the height of the prism.

The title question concerns the partition of I_d into the two sets

$$I_d^f = \{(i,j) \in I_d: \beta(d,i,j) < \infty\} \quad \text{and} \quad I_d^\infty = \{(i,j) \in I_d: \beta(d,i,j) = \infty\}.$$

Note that

(1) A member (i, j) of I_d belongs to I_d^∞ if and only if there is a sequence P_n $(n=1, 2, \cdots)$ of d-polytopes such that $\zeta_j(P_n)$ is bounded above while $\zeta_i(P_n)$ tends to infinity.

We close by showing

$$I_d^\infty \supset \{(i,j) \in I_d: i < j\} \tag{2}$$

and are tempted to conjecture that equality holds in (2), as it surely does when $d \leqq 4$. The validity of (2) for all d is an easy consequence of the fact that always

$$I_{d+1}^\infty \supset I_d^\infty \cup \{(i, d+1): (i-1, d) \in I_d^\infty\}. \tag{3}$$

For any d-polytope Q in E^d and any $\tau > 0$, let Q^τ denote the prism $Q \times [0, \tau]$, a $(d+1)$-polytope in E^{d+1}. From a description of the facial structure of Q^τ in terms of that of Q, it follows that

$$\zeta_k(Q^\tau) = 2\zeta_k(Q) + \tau\zeta_{k-1}(Q) \quad \text{for } d \geqq k \geqq 1, \quad \text{while} \tag{4a}$$

$$\zeta_{d+1}(Q^\tau) = \tau\zeta_d(Q). \tag{4b}$$

To see that $I_d^\infty \subset I_{d+1}^\infty$, consider an arbitrary $(i, j) \in I_d^\infty$, let P_n be as in (1), and use (4a) to see that if τ_n tends sufficiently rapidly to 0 then $\zeta_j(P_n^{\tau_n})$ is bounded above while $\zeta_i(P_n^{\tau_n})$ tends to ∞. To complete the proof of (3), suppose $(i-1, d) \in I_d^\infty$, let P_n be such that $\zeta_d(P_n)$ is bounded above while $\zeta_{i-1}(P_n)$ tends to ∞, and use (4a) and (4b) to see that if τ_n tends sufficiently rapidly to ∞ then $\zeta_{d-1}(P_n^{\tau_n})$ is bounded above while $\zeta_i(P_n^{\tau_n})$ tends to ∞.

References

1. O. Aberth, An isoperimetric inequality for polyhedra and its application to an extremal problem, Proc. London Math. Soc., (3) 13 (1963) 322–336.

2. H. G. Eggleston, B. Grünbaum, and V. Klee, Some semicontinuity theorems for convex polytopes and cell-complexes, Comment. Math. Helv., 39 (1964) 166–188.

3. D. G. Larman and P. Mani, Gleichungen und Ungleichungen für die Gerüste von konvexen Polytopen und Zellenkomplexen, Comment. Math. Helv., to appear.

4. Z. A. Melzak, Problems connected with convexity, Canad. Math. Bull., 8 (1965) 565–573.

CLASSROOM NOTES

EDITED BY DAVID DRASIN

Manuscripts for this Department should be sent to David Drasin, Division of Mathematical Sciences, Purdue University, Lafayette, IN 47907.

THE CONSTRUCTION OF SECOND ORDER DIFFERENTIAL EQUATIONS WITH SOLUTIONS HAVING GIVEN PROPERTIES

W. R. UTZ, University of Missouri

The most common problem of differential equations is that of determining properties of solutions of given equations. Here we wish to consider the converse problem, especially for second order equations. That is, given functions of a specified kind we shall seek second order equations within a given class for which the given functions are solutions. Clearly, the class of equations will be limited since there is, trivially, always an equation $y''=f''(x)$ satisfied by any given twice differentiable function $f(x)$.

Suppose that at first we require that a twice differentiable function satisfy an equation

$$y'' + p(x)y = 0 \tag{1}$$

for some continuous $p(x)$. That is, $y(x)$ is given and we seek a continuous $p(x)$ so that the given $y(x)$ satisfies (1) for this $p(x)$. One examines

$$\frac{-y''(x)}{y(x)} \tag{2}$$

and observes that $y(x)$ satisfies (1) if

$$p(x) = \frac{-y''(x)}{y(x)}$$

is, in fact, a continuous function. That is, if (2) has, at worst, removable singularities. Clearly, this is impossible unless $y''(x)$ vanishes when $y(x)$ vanishes. Thus, $y=\sin x$ satisfies an equation (1) for all x while $y=\sin x^2$ does not.

The function $y=\sin x^2$ (more generally, $y=A \sin x^2+B \cos x^2$) satisfies

$$xy'' - y' + 4x^3y = 0. \tag{3}$$

Thus all solutions of (3) are oscillatory and bounded.

If $\phi(x)\in C^2$, then $y=A \cos \phi(x)+B \sin \phi(x)$ satisfies

$$\phi'(x)y'' - \phi''(x)y' + (\phi'(x))^3y = 0 \tag{4}$$

and all solutions (4) are bounded.

If $\phi(x)\in C^2$, then $y=A \cosh \phi(x)+B \sinh \phi(x)$ satisfies

$$\phi'(x)y'' - \phi''(x)y' - (\phi'(x))^3y = 0. \tag{5}$$

Equations (3), (4), (5) illustrate the following theorem, the proof of which is immediate.

THEOREM. *If $y=y(x)$ satisfies $y''+b(x)y'+c(x)y=g(x)$ for all real x and if $u(x)\in C^2$ for all real x, then $y=y(u(x))$ satisfies*

$$u'(x)y'' - [u''(x) - b(u(x))(u'(x))^2]y' + c(u(x))(u'(x))^3y = (u'(x))^3g(u(x)).$$

This theorem, in special cases, is helpful in providing equations with solutions having a prescribed property. For example, to secure an equation having a solution with a prescribed number of zeros, let $\phi(x)=(2n+1)$ arc tan x in (4) to secure the equation

$$(1 + x^2)^2y'' + 2x(1 + x^2)y' + (2n + 1)^2y = 0. \tag{4'}$$

Solutions of (4′) include

$$y = \sin[(2n + 1) \text{ arc tan } x]$$

with an odd number, $2n+1$, of zeros and

$$y = \cos[(2n + 1) \text{ arc tan } x]$$

with an even number, $2n$, of zeros. All solutions of (4′) are bounded.

By appropriate choices of $\phi(x)$ in (4) and (5) one is able to provide equations with solutions having other specified properties. As further examples, solutions of (3) are bounded and oscillatory but not periodic; (5) with $\phi(x)=x \sin x$ has an unbounded, oscillatory solution $y=\sinh (x \sin x)$ and an unbounded nonoscillatory solution $y=\cosh(x \cos x)$.

All of the above examples are linear. We shall now consider the discovery of nonlinear equations with some solutions of a special kind.

The equation

$$x'' + \omega^2x = 0 \tag{6}$$

has the solutions $x=A \sin(\omega t+\beta)$, wherein A and β are arbitrary. John Brock [1] observes that since

$$x' = A\omega \cos(\omega t + \beta) \quad \text{and} \quad x'' = - A\omega^2 \sin(\omega t + \beta)$$

one has

$$\omega^2 = \frac{-x''}{x} \quad \text{and} \quad A^2 = x^2 - \frac{xx'^2}{x''}$$

independently of the differential equation (6). Hence if $f(u, v)$ is any function at all and ω^2 and A^2 are selected so that $f(\omega^2, A^2)=0$, then $x=A \sin(\omega t+\beta)$, for this choice of A and ω, are (β is arbitrary) solutions of the differential equation

$$f\left(\frac{-x''}{x}, x^2 - \frac{xx'^2}{x''}\right) = 0. \tag{7}$$

Thus, (7) provides a source of nonlinear equations permitting simple harmonic motion. Brock devotes particular attention to the special case

$$x'' + xg\left(x^2 - \frac{xx'^2}{x''}\right) = 0.$$

One may apply the device of Brock to other equations. For example,

$$x'' - \omega^2 x = 0 \tag{8}$$

is satisfied by $x = A\ \sinh(\omega t+\beta)$ wherein A and β are arbitrary. In this case, $\omega^2 = x''/x$, $A^2 = -x^2 + xx'^2/x''$. Thus, if A and ω are selected so that $f(\omega^2, A^2) = 0$, then $x = A\ \sinh(\omega t+\beta)$ satisfies

$$f\left(\frac{x''}{x}, \ -x^2 + \frac{xx'^2}{x''}\right) = 0.$$

As another example, the spring equations $x'' + ax \pm bx^3 = 0$ are satisfied by the Jacobian elliptic functions [2]. The method of Brock may be used to find other nonlinear equations satisfied by these functions.

In particular,

$$x'' + bx^3 = 0, \qquad b > 0 \tag{9}$$

(a hard spring) is satisfied by $x = A\ \mathrm{cn}(\omega t+\beta)$, where $k = \sin\alpha = 1/\sqrt{2}$ (the modulus of the elliptic functions to be used) and $\omega^2 = bA^2$. The A and β may be selected arbitrarily. Keeping in mind that $\sin\alpha = 1/\sqrt{2}$, we have

$$\begin{aligned}
x &= A\ \mathrm{cn}(\sqrt{b}\,|A|\,t+\beta) \\
x' &= -A\,|A|\,\sqrt{b}\,\mathrm{sn}(\sqrt{b}\,|A|\,t+\beta)\,\mathrm{dn}(\sqrt{b}\,|A|\,t+\beta) \\
x'' &= -A^3 b\ \mathrm{cn}^3(\sqrt{b}\,|A|\,t+\beta)
\end{aligned}$$

from which we have (independently of equation (9))

$$b = \frac{-x''}{x^3}, \qquad A^4 = x^4 - \frac{2x^3x'^2}{x''}.$$

Thus, if $f(u, v)$ is any function and b, A are selected so that $f(b, A^4) = 0$, then

$$f\left(\frac{-x''}{x^3}, \ x^4 - \frac{2x^3x'^2}{x''}\right) = 0$$

has, for this b and A, the two-parameter family of periodic solutions

$$x = A\ \mathrm{cn}(\sqrt{b}\,|A|\,t+\beta).$$

References

1. John E. Brock, Some non-linear systems permitting simple harmonic motion, J. Appl. Physics, 21 (1950) 238–243.

2. C. G. Lambe and C. J. Tranter, Differential Equations, Row, Peterson and Co.; Evanston, Ill., 1961.

A COUNTEREXAMPLE TO A LEMMA ON THE EXISTENCE OF SIMPLE REFINEMENTS

R. E. SMITHSON, University of Wyoming

Let X be a topological space, and let $x, y \in X$. Then a *simple chain* from x to y is a finite collection, $\{L_1, \cdots, L_k\}$ of subsets of X such that $x \in L_1$, $x \notin L_i$, $i \neq 1$, $y \in L_k$, $y \notin L_i$, $i \neq k$, and $L_i \cap L_j \neq \emptyset$ if and only if $|i-j| \leqq 1$. Then a simple chain $\mathcal{C} = \{C_1, \cdots, C_n\}$ is called a *simple refinement* (Hocking and Young [1] use the term "straight through" in this context) of a simple chain $\mathcal{L} = \{L_1, \cdots, L_k\}$ if

(1) for each $C_i \in \mathcal{C}$, there is an $L_j \in \mathcal{L}$ such that $C_i \subset L_j$, and

(2) if $C_i, C_j \subset L_s$, and if $i < t < j$, then $C_t \subset L_s$. Then in both [1] and [2] below the following lemma is stated. The lemma is then used to prove that every locally compact, connected, locally connected metric space is arcwise connected, and hence, this lemma is quite significant.

LEMMA. *Let $\mathcal{C}$ be a simple chain of connected open sets in X from x to y. If $\mathcal{U}$ is a collection of open sets such that each member of $\mathcal{C}$ is the union of members of $\mathcal{U}$, then there is a simple chain consisting of members of $\mathcal{U}$ from x to y which is a simple refinement of $\mathcal{C}$.*

NOTE. In the statement of this lemma in Hocking and Young [1, p. 116] there were some additional hypotheses given but they were not used in the alleged proof.

We now give a counterexample to the above lemma. All sets given below are subsets of the plane R^2 and the topology used is the relative topology.

Let $X_1 = \{(x, 0): -2 \leqq x \leqq 2\}$, $X_2 = \{(x, y): x^2 + y^2 = 1 \text{ and } y \geqq 0\}$ and set $X = X_1 \cup X_2$. Also let $Y_1 = \{(x, y) \in X_2 : x < 0 \text{ and } 0 \leqq y < \frac{1}{4}\}$ and $Y_2 = \{(x, y) \in X_2 : x > 0 \text{ and } 0 \leqq y < \frac{1}{4}\}$. Then let

$$C_1 = Y_1 \cup \{(x, 0): -2 \leqq x < \tfrac{3}{2}\} \cup Y_2$$

and

$$C_2 = \{(x, 0): -\tfrac{3}{2} < x < -\tfrac{1}{2}\} \cup X_2 \cup \{(x, 0): \tfrac{1}{2} < x \leqq 2\}.$$

Then C_1 and C_2 are open connected subsets of X and $\{C_1, C_2\}$ is a simple chain from $(-2, 0)$ to $(2, 0)$ in X. But if $\mathcal{U}$ consists of open subsets of X whose diameters are less than $\frac{1}{4}$, then condition (2) in the definition cannot be satisfied. This is because $C_1 \cap C_2$ is not connected and each component of $C_1 \cap C_2$ will contain a member of any simple chain from $\mathcal{U}$. But in chaining from one component of $C_1 \cap C_2$ to the other by members of $\mathcal{U}$ there will be at least one link either not contained in C_1 or not contained in C_2.

The difficulty in the "proof" given in [1] and [2] is that the authors did not consider the situation that occurs in the above example. To overcome this problem replace condition (2) above by the following condition. (This condition was suggested by the referee.)

(2'). If $C_i, C_j \subset L_s$ and if $i < t < j$, then $C_t \subset L_{s-1} \cup L_s \cup L_{s+1}$.

Then the proofs given in [1] and [2] are valid when (2) is replaced by (2′), and thus the lemma is true when conditions (1) and (2′) are used in the definition of a simple refinement, and the major theorem on arcwise connectedness can also be proved from conditions (1) and (2′).

References

1. J. G. Hocking and G. S. Young, Topology, Addison-Wesley, Reading, Mass., 1961.
2. T. O. Moore, Elementary General Topology, Prentice-Hall, Englewood Cliffs, N. J., 1964.

AN ELEMENTARY DISCUSSION OF THE TRANSCENDENTAL NATURE OF THE ELEMENTARY TRANSCENDENTAL FUNCTIONS

R. W. HAMMING, Bell Telephone Laboratories

When the elementary transcendental functions are introduced in the calculus course, it is usually stated that they are not algebraic functions, but little indication is given either as to what this means or how it can be proved. The purpose of this note is to fill this gap partially.

Probably the most convenient approach to this matter is the increasingly common path of introducing the $\ln x$ as

$$\ln x = \int_1^x dt/t,$$

which is equivalent to $d/dx(\ln x) = 1/x$; $\ln 1 = 0$.

We first prove that $\ln x$ is not a rational function, that is

$$\ln x \neq \frac{N(x)}{D(x)},$$

where $N(x)$ and $D(x)$ are polynomials with no common factor. If it were a rational function, then upon differentiating both sides of the equality we would have $1/x = (DN' - ND')/D^2$, or $D^2 = x(DN' - ND')$. We see that $D(x)$ has a factor x. Let

$$D(x) = x^k D_1(x), \quad D_1(0) \neq 0, \quad k \geqq 1.$$

Substituting and dividing out x^k, we get

$$x^k D_1^2 = xD_1N' - kND_1 - xND_1',$$

from which we see that $N(x)$ is divisible by x. The common factor of x in both $N(x)$ and $D(x)$ leads to a contradiction.

We are now ready to prove that

$$y = \ln x$$

is not algebraic, that is, there is no polynomial in x and y with real or complex

coefficients such that $f(x, y)=0$, or what is the same thing,

$$\sum_{k=0}^{N} P_k(x)y^k = 0, \qquad P_N(x) \not\equiv 0,$$

where the $P_k(x)$ are polynomials in x. There is an essentially unique equation of minimum degree N because if there were two equations of the same degree differing by more than a multiplicative factor then by eliminating the highest power of $\ln x$ between them we would have a lower degree equation. Assuming we have chosen the function $f(x, y)$ for which N is the smallest possible we can write the equation as

$$(\ln x)^N + \frac{P_{N-1}}{P_N}(\ln x)^{N-1} + \cdots + \frac{P_0}{P_N} = 0, \qquad N \geqq 2$$

and differentiate to get

$$N(\ln x)^{N-1} + x\left(\frac{P_{N-1}}{P_N}\right)'(\ln x)^{N-1} + \cdots = 0.$$

If all the terms in $(\ln x)^k$, $k=0, 1, \cdots, N-1$, do not vanish identically, then we have a lower degree polynomial (in $\ln x$), a contradiction. If all the terms do vanish, then in particular

$$\frac{N}{x} + \left(\frac{P_{N-1}}{P_N}\right)' = 0.$$

Integrating this, we find that $\ln x$ is a rational function, which we just proved is impossible. Hence, $\ln x$ is not an algebraic function.

There is a second idea of an algebraic function that the student needs to consider, namely that any finite combination of additions, subtractions, multiplications, divisions, and radicals with rational exponents of algebraic functions is still algebraic. In particular the student asks, "Is it possible that

$$\ln x = \frac{\sqrt{x^2+1} - \pi\sqrt[3]{x^2-1} + 2\sqrt[5]{x^2-2x+3} + x^2}{x^{1/3}\sqrt[7]{x+1} + 9\sqrt[11]{x^2+3} - (\sqrt[13]{x^2} + \pi^2)(\sqrt[7]{x^2+1})}$$

or something like it?" Note we are excluding x^π etc.

We now indicate the proof that this second definition is included in the first one. Consider the sum of two expressions, each of which is a root of a polynomial. Let this sum be

$$\alpha_1(x) + \beta_1(x),$$

where $\alpha_1(x)$ is a solution of $f_1(x, y)=0$ with the complete set of solutions

$$\alpha_1(x), \alpha_2(x), \cdots, \alpha_r(x).$$

Let $\beta_1(x)$ be a solution of $f_2(x, y)=0$ with the complete set of solutions

$$\beta_1(x), \beta_2(x), \cdots, \beta_s(x).$$

Now consider the set of rs functions $\alpha_i(x)+\beta_j(x)$ and the corresponding polynomial in y,

$$\prod_{i,j} (y - \alpha_i - \beta_j) = 0$$

having these rs factors as solution. This is a symmetric function in both the α_i and the β_j. We now use the theorem that every rational symmetric function is expressible rationally in terms of the elementary symmetric functions:

$$\begin{aligned} p_1 &= \alpha_1 + \alpha_2 + \cdots + \alpha_s \\ p_2 &= \alpha_1\alpha_2 + \alpha_1\alpha_3 + \cdots + \alpha_{s-1}\alpha_s \\ &\vdots \\ p_s &= \alpha_1\alpha_2\alpha_3 \cdots \alpha_s. \end{aligned}$$

But these are in turn rational expressions in the coefficients of $f_1(x, y)=0$. Similarly for the $\beta_j(x)$. Thus we have

$$\prod_{i,j} (y - \alpha_i - \beta_j)$$

as a rational expression in the coefficients of $f_1(x, y)=0$, $f_2(x, y)=0$ and assorted integers that arose in the algebraic manipulations are indicated.

Similar arguments show that differences, products, quotients, and radicals of algebraic expressions are again algebraic, and we have therefore shown that the second definition of an algebraic function is included in the first.

We now turn to the inverse function of ln x, namely e^x. Since ln x does not satisfy any polynomial

$$f(x, y) = 0$$

we have merely to set $x=e^t$ to get $f(e^t, t)=0$ as an equivalent impossibility.

For sin x, cos x, tan x, etc., if one of them, say sin x, satisfied

$$f(x, \sin x) = 0,$$

then the polynomial $f(x, 0)=0$ would have an infinite number of zeros, namely $x=0, \pm\pi, \pm 2\pi, \cdots$. Thus since the trigonometric functions have an infinite number of zeros they cannot be algebraic functions.

From the argument we used for the exponential function, we see that the corresponding inverse functions arcsin x, arccos x, arctan x, etc. also cannot be algebraic functions.

This presentation, *except* possibly that of the inclusion of the second definition of algebraic functions in the first, is readily presented in a calculus course to the better prepared students; the less prepared usually don't care, being willing to believe that the functions are transcendental (not algebraic).

Using similar methods, and slight extensions of them, integrals like

$$\int_0^x e^{-t^2}dt, \qquad \int_x^\infty (e^t/t)dt$$

can be shown to be transcendental. However, at this point in the development the real question is: Can these new functions be expressed as finite combinations of the algebraic and elementary transcendental functions we now have on hand? Unfortunately the above simple methods seem to be inadequate for this purpose. (See J. F. Ritt, Integration in Finite Terms, Columbia University Press, 1948, for a more sophisticated and more powerful treatment.)

Thanks for help in discussing this topic are due to A. J. Goldstein and Jessie MacWilliams.

MATHEMATICAL EDUCATION

EDITED BY J. G. HARVEY AND M. W. POWNALL

Material for this Department should be sent to either of the editors: J. G. Harvey, Department of Mathematics, University of Wisconsin, Madison, WI 53706; M. W. Pownall, Department of Mathematics, Colgate University, Hamilton, NY 13346.

ON THE MATHEMATICS CURRICULUM IN BLACK COLLEGES

A. J. SCAVELLA,† Tuskegee Institute

It is a pleasure and an honor to have been invited to address this august body of mathematicians on this occasion. When asked to participate, I was curious to know why. If in me Professor Stewart desired a first-rate mathematician, then I do not qualify; if in me he desired someone dedicated to the education of deprived youngsters, then perhaps I do qualify; if in me he desired someone black—then my Italian name and Jewish nose notwithstanding, I know I qualify!

On April 24–26, 1969, a Ford Foundation-sponsored Conference on Mathematics Curriculum was held at Morgan State College under the direction of Professor Walter Talbot, Chairman of the Department of Mathematics at that institution.

The purpose of the Conference was to discuss matters relating to curriculum, with a view toward improving the effectiveness of the individual mathematics programs at black colleges.

In attendance at the Conference were mathematics professors from twenty-seven colleges and universities with student bodies that are predominantly black—Afro-American, Negro, etc., and some resource persons, including Professors Richard Anderson and R. Creighton Buck as representatives from CUPM.

This Conference was much needed and very timely—for never before has there been as much concern about the status of the mathematics curriculum in black colleges as exists today. The Ford Foundation, in sponsoring the Con-

† We have just learned that Professor Scavella passed away on February 5. *Editor.*

A COMPARISON OF TWO APPROACHES TO TEACHING LIMIT THEORY

F. M. PAVLICK, Slippery Rock State College

Using the concept of a collection of advanced sets, E. J. McShane formulates one definition of limit, $\lim f(t)$, that is general enough to include all types of limits of real valued functions normally encountered in elementary calculus: $\lim_{x\to r} f(x)$, $\lim_{x\to\infty} f(x)$, $\lim_{j\to\infty} b_j$, [2]. The advanced sets definition can be applied to other concepts of the calculus which depend on the limiting process, e.g., $\int_a^b f(x)dx$ and $f'(x)$.

The traditional approach to limits is to introduce each limit definition separately. Little mention is made of the relationship between the various definitions. One wonders whether the use of the unifier, advanced sets, would provide a beginning calculus student with a better understanding of the underlying structure of the limit definitions by pointing out the properties common to each.

On the other hand, the addition of this concept to the limit definitions may be too much for a student to handle, especially at the time of his first formal introduction to limits. With this in mind, the success of the advanced sets approach may depend on the ability and achievement level of the individual calculus student, coupled with the question of generality versus specificity. What follows is a brief description of the author's doctoral dissertation which was designed to investigate this problem.

First, let us notice an important mathematical advantage of the all inclusive definition proposed by McShane. Once an appropriate fundamental theorem such as $\lim[f(t)+g(t)]=\lim f(t)+\lim g(t)$ has been proven in the general setting, it can be used in all the special cases without having to be proven over and over again. For example, one could conclude immediately that

$$\int_a^b [f(x)+g(x)]dx = \int_a^b f(x)dx + \int_a^b g(x)dx.$$

For a detailed discussion of this and other aspects of the advanced sets approach, see [2, 3].

Description of the study. The main problem in this study was to determine if there is a measurable difference in learning, as measured by an achievement test, between students who studied a traditional ϵ, δ presentation of limit theory (Treatment T) and students who studied an advanced sets approach to limits (Treatment A). Treatment T consisted of a programed unit which introduced the definitions, $\lim_{j\to\infty} b_j$ and $\lim_{x\to r} f(x)$, separately in the traditional manner. Treatment A was achieved by means of a program which used the advanced sets approach to limits.

Also included in the study was an attempt to determine if ability and achievement level would contribute to a difference in learning from these two treatments. To this end honors and nonhonors beginning calculus students at

The Florida State University were selected to represent two levels labeled H and N respectively.

The students in each level were randomly assigned to one of the two treatments which resulted in a 2×2 levels χ treatments design. The students read the programed units in four study sessions. A test was administered during the fifth session. This test was composed of two parts. Part I consisted of 23 multiple choice items primarily concerned with finding or recognizing limits of certain functions. Part II required the student to verify limits of some functions and to provide proofs of some simple theorems. A randomized 2×2 levels χ treatments analysis of variance was employed to test the following null hypotheses.

Null Hypotheses.

1. There is no significant difference between the mean scores of Level H and Level N on the criterion test.
2. There is no significant difference between the mean scores of Treatment Group A and Treatment Group T on the criterion test.
3. There is no interaction between levels (H and N) and treatment groups (A and T).

TABLE 1.—Scores on Section I of the criterion test

	Treatments						
	A			T			
	17	21		15	21		
	18	22	Mean 20.6	16	21	Mean 19.08	
H	19	22		16	21		Row Mean
	20	23	Std. Dev. 2.06	16	22	Std. Dev. 2.81	19.77
	21	23		17	22		
Levels				20	22		
	9	14		11	17		
	10	14	Mean 13.28	12	17	Mean 15.77	
	11	14		12	18		Row Mean
N	11	15	Std. Dev. 2.52	13	19	Std. Dev. 3.03	14.48
	12	16		14	19		
	13	16		16	20		
	13	18		17			
	Column Mean 16.33			Column Mean 17.36			

TABLE 2.—Analysis of variance of scores on Section I of the criterion test*

Source	Degrees of Freedom	Sum of Squares	Mean Square	F-ratio
Levels	1	340.58	340.58	48.12[a]
Treatments	1	2.82	2.82	0.40
Levels by Treatment	1	48.24	48.24	6.82[a]
Error	45	318.48	7.08	

[a] Significant at the 0.05 level.
* Obtained by means of computer program 05V [1].

Results and conclusions. Table 1 gives a summary of scores by cell, treatment, and level for Part I of the criterion test. Table 2 is the source table for the analysis of variance performed on these scores. Since similar results were obtained for Part II of the test, these data are not included in this paper.

Significant differences between levels results in rejection of hypothesis 1. This is not surprising since the honors class consists of select students. Significant interaction results in rejection of hypothesis 3 which suggests that the effectiveness of a particular treatment depends on the level of the individual student. The profile (Figure 1) indicates that the traditional approach may be favored by non-honors students and the advanced set approach by honors students.

The t-statistic was used to compare cell means. A $t > 2.06$ was required for an indication of significant difference with 95% confidence. The mean scores for the Level N students studying Treatments A and T were 13.28 and 15.77 respectively. When these two means were compared, a $t = 2.32$ was obtained. A Mann-Whitney U Test further confirmed this significant difference. Thus, those nonhonors students who studied the traditional approach performed significantly better on the criterion test than those who studied the advanced set approach.

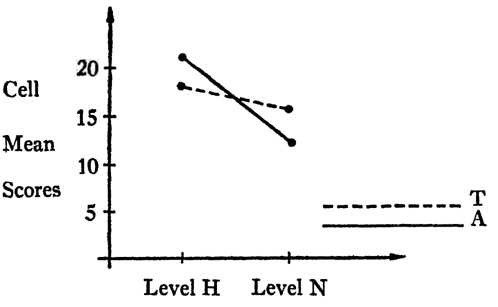


FIG. 1. Profiles for the cell means of the scores on Section I of the criterion test.

One must keep in mind that this was a short term study performed with only 49 students, but the results in general lead one to believe that the efficiency

of the advanced set approach to limits depends on the ability and achievement level of the calculus student. Students at a high level when taught by the advanced set approach should do at least as well as when taught by the traditional approach. However, students of average achievement would profit more from the traditional strategy of instruction.

Discussion. One possible explanation for these results may be found in whole-part vs. part-whole learning theory. The traditional approach follows a "by parts" strategy whereas the advanced set approach has many characteristics of a "whole-part" teaching procedure. Seagoe and Symonds [4, 5] theorize a reciprocal relationship between the ability of the learner and the difficulty of the subject matter. With subject matter of given difficulty the more advanced student can learn more effectively by the whole-part method of instruction. With students of given ability the easier subject matter can be learned more readily by the whole-part method. Evidently, the nonhonors students were able to comprehend the limit concept presented in parts ($\lim_{j\to\infty} b_j$ and $\lim_{x\to r} f(x)$), whereas they were less able to grasp the significance of the concept as a whole (the all inclusive definition of limit presented in the advanced set approach). The subject matter was too difficult for these students to learn it more effectively by the whole method. For the honors students, on the other hand, it was at a level of difficulty at which they could actually profit from studying the more general approach.

To the author's knowledge, no calculus text defines a collection of advanced sets. However, many contemporary texts do incorporate the neighborhood concept into their limit definitions. Although the neighborhood approach is not so general as the advanced sets approach, it does require learning of additional terminology. Such a text could not be read with profit by (average) students at Slippery Rock. One utilizing the ϵ, δ approach is now being used with much better results. Thus, what was suggested by the results of the study, seems to be applicable in practice. Evidently a calculus text should not be rated unconditionally as "good" but rather "good for whom."

References

1. General Linear Hypothesis, BMDO5V, Biomedical Computer Programs, rev. ed., edited by W. J. Dixon, pp. 543–557, Health Sciences Computing Facility, Department of Preventive Medicine and Public Health, School of Medicine, University of California, Los Angeles, 1965.

2. E. J. McShane, Theory of limits, MAA Film Manual #2, Mathematical Association of America, 1962.

3. ———, A Theory of Limits, Studies in Modern Analysis, vol. 1, Studies in Mathematics, ed. by R. C. Buck, Mathematical Association of America, 1962.

4. M. V. Seagoe, Qualitative wholes: a re-evaluation of the whole-part problem, J. Ed. Psychol., 27 (1936) 537–545.

5. P. M. Symonds, What education has to learn from psychology, Teachers College Record, 58 (1957) 329–339.

PROBLEMS AND SOLUTIONS

EDITED BY EMORY P. STARKE

All problems (both elementary and advanced) proposed for inclusion in this Department should be sent to E. P. Starke, 1000 Kensington Ave., Plainfield, N. J. 07060. Proposers of problems are urged to enclose any solutions or information that will assist the editors. Ordinarily, problems in well-known textbooks and results in generally accessible sources are not appropriate for this Department. No solutions (except those accompanying proposals) should be sent to Professor Starke.

ELEMENTARY PROBLEMS

Solutions of Elementary Problems should be sent to Problems Group, Mathematics Department, University of Maine, Orono, ME 04473. To facilitate their consideration, solutions of Elementary Problems in this issue should be typed (with double spacing) and should be mailed before June 30, 1970. Contributors (in the United States) who desire acknowledgment of receipt of their solutions are asked to enclose self-addressed stamped postcards.

E 2221. *Proposed by Ioan Tomescu, University of Bucharest, Rumania*

Let ABC and $A'B'C'$ be two triangles having sides a, b, c and a', b', c' respectively. If R is the circumradius of ABC, prove the inequality:

$$\frac{a^2}{a'} + \frac{b^2}{b'} + \frac{c^2}{c'} \leqq R^2 \frac{(a' + b' + c')^2}{a'b'c'}.$$

When does equality hold?

E 2222. *Proposed by R. M. Krause, National Science Foundation*

The following problem has been making the rounds at Berkeley, M. I. T., etc. Let N be any number between 0001 and 9998, excluding only 1111, 2222, etc. Arrange the digits in ascending order of magnitude and in descending order of magnitude and take the difference of the two resulting numbers. Call this difference $T(N)$. Show that repeated applications of the operator T converge on the number 6174. Is generalization possible?

E 2223. *Proposed by M. Slater, University of Bristol, England*

A pack of cards is shuffled by starting with the first card and then placing successive cards alternately above and below the growing discard pile. After repeated shuffles of this type, will the pack first return to its original order when the original top card first returns to its top position?

E 2224. *Proposed by D. P. Giesy, University of Southern California*

Suppose $\{a_n\}$ is a sequence of nonnegative real numbers such that $\limsup_{n\to\infty} (a_1+ \cdots +a_n)/n < \infty$ and $\lim_{n\to\infty} a_n/n=0$. Does it necessarily follow that $\lim_{n\to\infty} (a_1^2+ \cdots +a_n^2)/n^2=0$?

E 2225. *Proposed by Diane Comer and J. J. Le Tourneau, Fisk University*

Show that any positive integer S can be written in exactly k different ways as the sum of two or more consecutive positive integers (in increasing order), where k is the number of positive odd divisors of S greater than 1.

E 2226. *Proposed by M. S. Klamkin, Ford Scientific Laboratory*

If one altitude of a tetrahedron intersects two other altitudes, then all four altitudes are concurrent.

E 2227. *Proposed by N. S. Mendelsohn, University of Manitoba*

Find the greatest common divisor of

$$\binom{2n}{1}, \binom{2n}{3}, \binom{2n}{5}, \ldots, \binom{2n}{2n-1}.$$

SOLUTIONS OF ELEMENTARY PROBLEMS

A Sum of Stirling Character

E 2159 [1969, 300; corrected 1969, 690]. *Proposed by G. M. Lee, San Mateo, California*

Prove:

$$\sum_{r=1}^{n} (-1)^{r-1} \frac{1}{r} \sum_{i=0}^{r} (-1)^{r-i} \binom{r}{i} i^n = 0,$$

where $n=2, 3, 4, \cdots$.

I. *Solution by G. C. Dodds, State College, Pa.* The given sum may be rewritten:

$$\sum_{r=1}^{n} \sum_{i=0}^{r} (-1)^{2r-i-1} \frac{1}{r} \binom{r}{i} i^n = \sum_{r=1}^{n} \sum_{i=0}^{r} (-1)^{i+1} \frac{i}{r} \binom{r}{i} i^{n-1}.$$

By further rearrangement, and discarding the term for $i=0$,

$$\sum_{i=1}^{n} (-1)^{i+1} i^{n-1} \sum_{r=i}^{n} \frac{i}{r} \binom{r}{i} = \sum_{i=1}^{n} (-1)^{i+1} i^{n-1} \sum_{r=i}^{n} \binom{r-1}{i-1}.$$

Since

$$\sum_{r=i}^{n} \binom{r-1}{i-1} = \binom{n}{i},$$

the given sum is

$$\sum_{i=1}^{n}(-1)^{i+1}i^{n-1}\binom{n}{i}. \tag{1}$$

Now,

$$(1-e^x)^n=\sum_{i=0}^{n}\binom{n}{i}(-1)^i e^{(n-i)x}.$$

Hence, the $(n-1)$th derivative of this expression, evaluated at $x=0$, is (1) above. But this derivative at $x=0$ is clearly zero for $n=2, 3, \cdots$, so that the desired sum is zero for $n=2, 3, \cdots$.

II. *Solution by John Riordan, Rockefeller University.* The inner sum is $\Delta^r 0^n$ or $r!S(n, r)$, the Stirling number of the second kind, and $(-1)^{r-1}(r-1)!$ is $s(r, 1)$, the Stirling number of the first kind. Hence the result is the instance $m=1$ of the familiar orthogonal relation

$$\sum_{r=m}^{n} S(n, r)s(r, m) = \delta_{nm},$$

the Kronecker delta; and $\delta_{n1}=0$ for $n\neq 1$.

Also solved by Günter Bach (Germany), Anders Bager (Denmark), M. T. Bird, Michael Bousquet, C. A. Church, Jr., N. Ersec, A. F. Gentzel, F. Göbel & Tj. Plomp (Netherlands), M. G. Greening (Australia), J. C. Hickman, R. L. Jow, D. G. Kabe, J. Kaucký (Czechoslovakia), Harry Lass, Douglas Lind (England), O. P. Lossers (Netherlands), W. Moser, C. B. A. Peck, Richard Post, E. A. Power (Australia), M. L. Railkar (India), Simeon Reich (Israel), John Riordan, M. J. Rossin, E. F. Schmeichel, R. R. Seeber, C. V. L. Smith, F. C. Smith, DB²S, Stephen Spindler, Dragutin Surtan (Yugoslavia), T. Tamura (Japan), M. R. Wise, Chung-kiu Wong, and the proposer.

Robert Breusch, L. Carlitz and Alexander Zujus solved assumed reformations of the problem other than the proposer's corrected version. Earlier, Bager had proposed essentially this problem in *Elemente der Mathematik*; see solutions by O. Reutter, I. Paasche and L. Carlitz, 1963, 5.

Quadratic Nonresidues Summing to Zero

E 2173 [1969, 553]. *Proposed by Emanuel Vegh, Naval Research Laboratory*

If p is a prime ($p\neq 2$, 3, 5, 11, or 17) there are three distinct quadratic nonresidues of p, whose sum is divisible by p.

Solution by Joel Spencer, University of Maine. Since $7\mid 3+5+6$ and $13\mid 2+5+6$, we assume $p\geqq 19$. Let g be a fixed nonresidue (mod p). Then $a+b+c\equiv 0$ (mod p) where a, b, c are distinct nonresidues if and only if $ga+gb+gc\equiv 0$ (mod p) where ga, gb, gc are distinct residues. We solve the latter equation.

CASE I. $(-1\mid p)=-1$.

Subcase A: If $(2\mid p)=-1$, then $(8\mid p)=-1$; so for one of $n=4$, 5, 6, or 7,

$$\left(\frac{n}{p}\right)=+1, \quad \left(\frac{n+1}{p}\right)=-1,$$

giving $1+n+(-1-n)\equiv 0 \pmod{p}$ where $1, n, -1-n$ are distinct residues.

Subcase B: If $(2|p)=+1$, let n be the smallest integer such that

$$\left(\frac{n+1}{p}\right) = -1.$$

Since there are $\frac{1}{2}(p-1)$ nonzero residues, $n \leqq \frac{1}{2}(p-1)$. Since the residues are closed under multiplication, $n < \frac{1}{2}(p-1)$; so $1, n, -1-n$ are distinct residues whose sum is $0 \pmod{p}$.

CASE II. $(-1|p)=+1$.

If $(5|p)=+1$, then $1+4+(-5)=0$ and $1, 4, -5$ are distinct nonresidues. If

$$\left(\frac{10}{p}\right) = +1,$$

then $1+9+(-10)=0$. Let

$$\left(\frac{5}{p}\right) = \left(\frac{10}{p}\right) = -1.$$

Then $(2|p)=+1$; so $(8|p)=+1$ and $1+8+(-9)=0$.

Also solved by Anders Bager (Denmark), D. M. Bloom, G. C. Dodds, M. G. Greening (Australia), Erwin Just, Ekkehard Kroll (Germany), D. C. B. Marsh, K. A. Ribet, E. W. Trost (Switzerland), K. M. Wilke, and the proposer.

Kroll proves that each nonresidue is a member of such a triple, while Trost shows that the number of such triples is

$$[(p-1)/48]\{p-8-3(-1|p)[1+2(2|p)]\}.$$

A Functional Equation

E 2176 [1969, 554]. *Proposed by R. S. Luthar, University of Wisconsin at Waukesha*

Find all continuous real functions f such that

$$(1) \qquad f\left(\frac{x+y}{x-y}\right) = \frac{f(x)+f(y)}{f(x)-f(y)}.$$

Solution by L. E. Ward, Sr., Escondido, California. Placing $y=0$ shows that $f(1)[f(x)-f(0)]=f(x)+f(0)$, whence we infer that $f(0)=0$ and $f(1)=1$, since $f(x)$ cannot be a constant. Replacing y by $x-2$ and again x by $x-1$ and y by 1, we obtain

$$(2) \qquad f(x-1) = \frac{f(x)+f(x-2)}{f(x)-f(x-2)}, \qquad f\left(\frac{x}{x-2}\right) = \frac{f(x-1)+1}{f(x-1)-1}.$$

It therefore follows that

$$f\left(\frac{x}{x-2}\right)=\frac{f(x)}{f(x-2)}\,. \tag{3}$$

From (3), with $x=4$, we have $f^2(2)=2f(2)$. But $f(2)\neq 0$, since, from (1) $f(2)=0$ would imply $f((x+2)/(x-2))\equiv 1$. Hence $f(2)=2$.

Now assume that $f(x)=x$ for $x=0, 1, 2, \cdots, n$ $(n\geqq 2)$. Then, with $x=n+1$, the first equation of (2) gives at once $f(n+1)=n+1$. It follows that $f(x)=x$ for all nonnegative integers.

Putting $y=cx$ yields

$$f\left(\frac{1+c}{1-c}\right)=\frac{f(x)+f(cx)}{f(x)-f(cx)}=\frac{1+f(c)}{1-f(c)},$$

from which $f(cx)=f(c)f(x)$. Now let $x=p/q$, a rational fraction, and $c=q$. Then $f(p)=f(p/q)f(q)$, showing that $f(x)=x$ for all positive rational numbers. By continuity it follows that $f(x)=x$ for all positive numbers.

Finally, placing $y=-x$ in (1) gives $f(0)=f(x)+f(-x)$, so that $f(-x)=-f(x)=-x$. We conclude that $f(x)\equiv x$.

Also solved by Joel Anderson, Anders Bager (Denmark), Joseph Beer, Stephen Berman, M. T. Bird, Michel Bousquet, Ted Cullen, R. J. Egbert, W. F. Fox, George Gastl & V. Sehgal, Michael Goldberg, M. G. Greening (Australia), Emil Grosswald, D. W. Hadwin, Erhard Heil, D. A. Herrero, Toshi Iida (Japan), B. G. Klein, Robert Kopp, Beatriz Margolis (Argentina), J. V. Michalowicz, Steven Minsker, Jernej Polajnar (Yugoslavia), Simeon Reich (Israel), E. F. Schmeichel, Hanna Schwerdtfeger, Hans Schwerdtfeger, Stewart Shapiro, R. A. Struble, Charles Wexler, Albert White, and the proposer.

Note. Several solutions attempting to prove by induction that $f(n)=n$ for positive integers n, were incomplete: they did not show that $f(2)=2$.

A Characteristic Value Problem

E 2177 [1969, 554]. *Proposed by P. M. Gibson, University of Alabama, Huntsville*

Let r be a nonzero complex number, and m an integer. Suppose that $A=(a_{ij})$, $B=(b_{ij})$ are $n\times n$ complex matrices with $b_{ij}=r^{m+i-j}a_{ij}$ for $i, j=1, \cdots, n$. Show that if λ is a characteristic value of A of multiplicity k then $r^m\lambda$ is a characteristic value of B of multiplicity k. Use this to show that if A is a tridiagonal matrix of odd order with zero diagonal then A is singular.

Solution by J. R. Kuttler, The Johns Hopkins University Applied Physics Laboratory. We have $B=r^mDAD^{-1}$, where D is the diagonal matrix with $d_{ii}=r^i$, $i=1, \cdots, m$. Since DAD^{-1} is a similarity transformation, the eigenvalues of B are r^m times the eigenvalues of A. Now let $m=0$, $r=-1$. When A has zero diagonal, $B=-A$. Thus A and $-A$ have the same eigenvalues. Hence the eigenvalues of A are symmetrically distributed about the origin and the number of nonzero eigenvalues is necessarily even. Thus, when n is odd there must be a zero eigenvalue, i.e., A is singular (cf. Problem 5488, [1967, 425]).

Also solved by J. V. Michalowicz, Simeon Reich (Israel), W. A. Thrash, Jr., and the proposer.

Continuous Real-valued Functions with at most one Nonzero Partial Derivative

E 2178 [1969, 690]. *Proposed by W. G. Dotson, North Carolina State University*

(1) Find a connected open set D in the plane, and a continuous function f from D to the real numbers, such that $f_x(x, y)\equiv 0$ on D but f is not a function of y alone. (2) Find (weakest possible) conditions on a connected open set D in R^n which insure that if f is a continuous function from D to the real numbers, and $\partial f/\partial x_i \equiv 0$ for $i=2, 3, \cdots, n$, then f is a function of x_1 alone.

Solution by Simeon Reich, Israel Institute of Technology. (1) Let D be the following set:

$$\{(x, y) \mid -2 < x < 2, 0 < y < 2\} \backslash \{(x, y) \mid -1 \leqq x \leqq 1, 1 \leqq y < 2\},$$

and let $f(x, y)$ be defined as follows:

$$f(x, y) = \begin{cases} -y + 2, & -2 < x < -1,\ 1 \leqq y < 2 \\ y & \text{for other } (x, y) \in D. \end{cases}$$

Here, f is not a function of y alone because $1/2 = f(-3/2, 3/2) \neq f(3/2, 3/2) = 3/2$.

(2) The following condition is sufficient: For any two distinct points $(x_1, x_2, \cdots, x_n)$, $(y_1, y_2, \cdots, y_n)$ belonging to D such that $x_i = y_i$ for $i=1, \cdots, j-1, j+1, \cdots, n$, where $j \geqq 2$, the set

$$\{(x_1, \cdots, z_j, \cdots, x_n) \mid x_j \leqq z_j \leqq y_j \quad \text{or } y_j \leqq z_j \leqq x_j\}$$

is a subset of D.

Proof. Take any two points $a = (x_1, x_2, \cdots, x_n)$, $b = (x_1, y_2, \cdots, y_n)$ belonging to D. We have to show that $f(a) = f(b)$. Indeed we have

$$\begin{aligned} f(a) - f(b) = \\ [f(x_1, x_2, \cdots, x_n) - f(x_1, y_2, x_3, \cdots, x_n)] + [f(x_1, y_2, x_3, \cdots, x_n) \\ - f(x_1, y_2, y_3, x_4, \cdots, x_n)] + \cdots + [f(x_1, y_2, \cdots, y_{n-1}, x_n) \\ - f(x_1, y_2, \cdots, y_n)]. \end{aligned}$$

Consider, for example, the first bracket. The function $F(z) = f(x_1, z, x_3, \cdots, x_n)$ is differentiable for $x_2 \leqq z \leqq y_2$ (or for $y_2 \leqq z \leqq x_2$, as the case may be). Therefore there is a number w, $x_2 < w < y_2$, such that $F(x_2) - F(y_2) = (x_2 - y_2)F'(w)$. But $F'(w) = f_{x_2}(x_1, w, x_3, \cdots, x_n) = 0$. Similarly for the other brackets. The result follows.

I do not know if this condition is the weakest possible in the sense that for any D not satisfying this requirement there exists a suitable function for which the conclusion is not valid. But note that neither the continuity of f nor the connectedness of D was needed in the proof.

Also solved by G. A. Heuer, E. F. Schmeichel, Luis Verde-Star, and the proposer.

ADVANCED PROBLEMS

All solutions of Advanced Problems should be sent to J. Barlaz, Rutgers—The State University, New Brunswick, N. J. 08903. Solutions of Advanced Problems in this issue should be typed (with double spacing) on separate, signed sheets and should be mailed before June 30, 1970. Contributors (in the United States) who desire acknowledgment of receipt of their solutions are asked to enclose self-addressed stamped postcards.

An asterisk (*) means neither the proposer nor the editors supplied a solution.

5720. *Proposed by Otto Morphy, Atlantis University*

Using the definition of "normal" not containing "Hausdorff," find all topological spaces X which are not normal but are such that each proper subspace is normal.

5721. *Proposed by Simeon Reich, The Technion, Haifa, Israel*

Let f: $K^n \to E^n$ be a continuous function where E^n is Euclidean n-space and $K^n = \{x \mid x \in E^n,\ \|x\| \leq 1\}$ such that for every $y \in S^{n-1} = \{x \mid x \in E^n,\ \|x\| = 1\}$ there is no $m > 1$ with $f(y) = my$. Show that f has a fixed point.

5722. *Proposed by D. G. Cantor, University of California, Los Angeles*

Let X be a compact subset of the reals. Prove that the necessary and sufficient condition for the existence of a nonconstant monic polynomial with real coefficients which has absolute value <1 on X is that there exist such a polynomial which has absolute value <2 on X. Show that "2" is sharp; i.e., 2 cannot be replaced by any larger number.

5723*. *Proposed by H. D. Ruderman, Hunter College High School*

Let $M = ab + cd - ef$. A random integer is selected uniformly from the numbers 0 to 9 inclusive and is used for some one of the six variables a, b, c, d, e, or f. Then a second random integer is selected independently from the same uniform distribution and is used for some other of the remaining variables. This selection and assignment is continued in the same way until six integers have been selected and assigned to the six variables. What is the best strategy for assigning the integers so as to maximize the expectation of M?

5724*. *Proposed by Paul Erdös, University College of Swansea, Wales*

Denote by $P(n)$ the largest prime factor of n. Let $f(n)$ be an increasing function of n so that $\sum_{n=1}^{\infty} 1/nf(n) < \infty$. Prove that $\sum_{n=1}^{\infty} 1/nf(P(n)) < \infty$.

5725. *Proposed by D. J. Lutzer, University of Washington*

Is it true that every compact Hausdorff space is a compactification of a metric space?

5726*. *Proposed by D. P. Allen, Jr., Bell Telephone Laboratories, Holmdel, N. J.*

Let X be a finite nonempty set, let ΣX $[\Delta X]$ denote the free [free commutative] semigroup on X, and let ϕ: $\Sigma X \to \Delta X$ denote the natural homomorphism. Do there exist semigroups T and T' of ΣX with $T \cup T' = \Sigma X$ and $T \cap T' = \emptyset$

such that (i) $\phi(T)\cap\phi(T')\neq\varnothing$ and (ii) neither T nor T' contains either a right or a left ideal of ΣX?

SOLUTIONS OF ADVANCED PROBLEMS

Rational Distances on a Conic Section

5650 [1969, 95]. *Proposed by D. E. Daykin, University of Malaya, Kuala Lumpur*

Let S be a set of points of the real x,y plane such that the distance between any two of the points is rational. How big can S be when S lies (1) on a hyperbola, (2) on a parabola, (3) on a proper ellipse, and (4) when S is of the form $(x_1, +d)$, $(x_1, -d)$, $(x_2, +d)$, $(x_2, -d)$, $\cdots$?

Note and comments by the proposer. It is known, and easy to verify, that there can be infinitely many points on a circle for which all distances are rational—choose the points $(1, \theta)$ for which $\tan\frac{1}{4}\theta$ is rational. By problem 5330 [1966, 1020] the set must be countable. S can consist of two points on one line and infinitely many points on another line—see D. E. Daykin, *Rational Polygons*, Mathematika, 10 (1963), 125–131. See also the reference in 5330.

Cases (1) *and* (3). We require $(x_i-x_j)^2+(y_i-y_j)^2$ equal to a square number, subject to $(x/a)^2\pm(y/b)^2=1$ for the ellipse and hyperbola respectively.

Case (4). We need a set of points x_i such that $(x_i-x_j)^2+4d^2$ is a square. We can assume that $x_1=0$ and, by a change of scale, that $4d^2=1$. Thus the problem is equivalent to finding a subset A of the set P of solutions of the Pell equation x^2+1 equals a square, such that $0\in A$ and $x_1, x_2\in A$ implies $(x_1-x_2)\in P$.

Case (2). Take for the parabola $y=at^2+bt+c$, $a\neq 0$. Putting $t=x-(b/2a)$ gives $y=ax^2+d$ for some d. We need a set of points (x_i, y_i) such that $y_i=ax_i^2+d$ and

$$(x_i - x_j)^2 + (y_i - y_j)^2 = \square,$$

which is, upon substituting for y_i, y_j,

$$(x_i - x_j)^2 + (ax_i^2 - ax_j^2)^2 = \square,$$

that is

$$1 + (ax_i + ax_j)^2 = \square.$$

So, if P is the set of solutions of the Pell equation $x^2+1=\square$, the problem is equivalent to finding a set B of numbers, such that $x_1, x_2\in B$ implies $x_1+x_2\in P$.

Note. The editors will be pleased to print partial solutions if these are available.

A Trigonometric Equation

5661 [1969, 309]. *Proposed by G. J. Foschini, Bell Telephone Laboratories, Holmdel, N. J.*

Find all solutions in the complex plane of

$$z = \sum_{q=2}^{\infty} \sum_{p=1}^{q} e^{2\pi ipz/q}.$$

Solution by Douglas Campbell, University of North Carolina. For the given double sum even to make sense we must have

$$\lim_{q\to\infty} \left| \sum_{p=1}^{q} e^{2\pi ipz/q} \right| = 0.$$

If z is not an integer·

$$\lim_{q\to\infty} \left| \sum_{p=1}^{q} e^{2\pi ipz/q} \right| = \lim_{q\to\infty} \left| e^{2\pi iz/q} \left(\frac{1 - e^{2\pi iz}}{1 - e^{2\pi iz/q}} \right) \right| = \infty.$$

Hence, the only possible solutions are integers. Since

$$\sum_{p=1}^{q} e^{2\pi ipn/q} = \begin{cases} 0 & \text{if } q \nmid n \\ q & \text{if } q \mid n, \end{cases}$$

we need only find all integral solutions to the equation

$$\sum_{q=2}^{\infty} \sum_{p=1}^{q} e^{2\pi ipn/q} = \sum_{\substack{q \mid n \\ q \geq 2}} q = n.$$

This latter equation is clearly true if and only if n is a positive prime. Therefore the only solutions to the original equation are the positive primes.

REMARK: The above analysis shows that the set of solutions of the equation $2z = \sum_{q=1}^{\infty} \sum_{p=1}^{q} e^{2\pi ipz/q}$ is the set of perfect numbers.

Also solved by T. M. Apostol, Robert Breusch, L. Carlitz, Josef Daneš (Czechoslovakia), Emil Grosswald, H. Pétard, Jernej Polajnar (Yugoslavia), Steve Rohde, and the proposer.

Inherited Property for a Polynomial Ring

5662 [1969, 309]. *Proposed by Irving Kaplansky, University of Chicago*

Let R be an associative ring with unit. Suppose in R whenever $ab = 1$ then $ba = 1$. Prove that the same thing holds in $R[x]$, the polynomial ring over R.

Solution by Sue Ann Hemr, Champaign, Illinois. Let $p(x)q(x) = 1$ and $q(x)p(x) = e(x)$. From the equation $p(0)q(0) = 1$ it follows that $q(0)p(0) = 1 = e(0)$. If $e(x) \neq 1$ then there is a minimal natural number m and a nonzero element r in R such that $e(x) = 1 + rx^m + \cdots$. From the relation between $p(x)$ and $q(x)$ we obtain $e(x) = e(x)e(x) = 1 + 2rx^m + \cdots$ and thus $r = 2r$. This contradicts the assumption that $r \neq 0$ and so $e(x) = 1$ after all.

Also solved by forty-four other readers.

E. J. Taft observes that the property remains true for the formal power-series ring $R[[x]]$ if the element 1 is replaced by any central element c which is not a zero divisor in R, i.e., $ab = c$ implies $ba = c$ becomes the hypothesis. He then raises the question whether $R[x]$ retains the property if we delete the assumption that c is a central nonzero divisor in R.

THE AMERICAN
MATHEMATICAL MONTHLY

(FOUNDED IN 1894 BY BENJAMIN F. FINKEL)

THE OFFICIAL JOURNAL OF

THE MATHEMATICAL ASSOCIATION OF AMERICA, INC.

VOLUME 77 NUMBER 4

CONTENTS

(continued on inside cover)

APRIL 1970

NOTICE TO AUTHORS

Articles should be typewritten on good quality paper, triple spaced with wide margins. Submit the original and a duplicate; keep a complete copy for protection against loss. Please follow the format used in current issues of the MONTHLY. The manuscript must make the author's intent clear to the printer. A paper accepted for publication by the editor but not acceptable to the printer will be returned to the author for retyping. Please use the American Mathematical Society *Manual for Authors of Mathematical Papers* as your guide and please conform to correct English usage. A *Manual for Monthly Authors* is available on request.

Prospective MONTHLY authors are advised to consult the Statement of Policy in the January 1969 issue, p. 2. Specialized research is usually unsuitable. Backlog: Main Articles 11 months, Math. Notes 10 months, Research Problems 6 months, Classroom Notes 7 months, Math. Education 6 months.

EDITORIAL CORRESPONDENCE: to HARLEY FLANDERS, American Mathematical Monthly, Purdue University, Lafayette, IN 47907; ADVERTISING CORRESPONDENCE: to RAOUL HAILPERN, Mathematical Association of America, SUNY at Buffalo, N. Y. 14214; CHANGE OF ADDRESS and SUBSCRIPTIONS: to A. B. WILLCOX, Mathematical Association of America, 1225 Connecticut Ave., N.W., Washington, D.C. 20036.

Annual dues for members of the Association (including a subscription to the American Mathematical Monthly) are $10.00. For nonmembers the subscription price is $15.00.

PUBLISHED BY THE ASSOCIATION at Washington, D. C., and Menasha, Wisconsin during the months of January, February, March, April, May, June–July, August–September, October, November, December.

Second-class postage paid at Washington, D. C., and additional mailing offices. Acceptance for mailing at special rate of postage provided for in the Act of February 28, 1925, embodied in Paragraph 4, Section 538, P.L. and R., authorized April 1, 1926.

PRINTED IN THE UNITED STATES OF AMERICA

OPTIMAL STOPPING

HERBERT ROBBINS, Columbia University

The theory of probability began with efforts to calculate the odds in games of chance. In this context, optimal stopping problems concern the effect on a gambler's fortune of various possible systems for deciding when to stop playing a sequence of games. Such problems are of interest in statistics, where the experimenter must constantly ask whether the increase in information contained in further data will outweigh the cost of collecting it.

Optimal stopping theory provides a general mathematical framework in which such problems can be precisely formulated and in some cases solved completely. The examples which we shall consider here are of a simpler nature than those arising in statistics, but will serve to illustrate some of the problems that arise in the general theory.

EXAMPLE 1. A fair coin is tossed repeatedly. After each toss we have the option of stopping or going on to the next toss, our decision at each stage being allowed to depend on the outcome thus far. We must stop after some finite (but not necessarily preassigned) number of tosses, and it is agreed that if we stop after the nth toss we are to receive a reward x_n which is a given function of the outcomes of the first n tosses. When should we stop so as to maximize our expected reward?

Let us introduce random variables $y_1, y_2, \cdots$ to represent the successive tosses, the y_i being independent with the common probability distribution $P(y_i=1)=P(y_i=-1)=1/2$; $y_i=1$ denoting heads on the ith toss and $y_i=-1$ tails. The reward sequence will then consist of a sequence of functions $x_1, x_2, \cdots$, where $x_n=f_n(y_1, \cdots, y_n)$. A *stopping rule* is then a random variable t with values in the set $\{1, 2, 3, \cdots\}$ and such that the event $[t=n]$ depends solely on the values of $y_1, \cdots, y_n$ and not on future values $y_{n+1}, \cdots$. Using a stopping rule t our reward x_t will be a random variable whose expectation Ex_t measures the performance on the average of the stopping rule t. The supremum $V=\sup\{Ex_t\}$ over the class C of all possible stopping rules t for which Ex_t exists is called the *value* of the sequence $\{x_n\}$, and if a stopping rule t exists such that $Ex_t=V$, t is said to be *optimal.*

Prof. Robbins earned his Harvard Ph.D. under Hassler Whitney. He followed an assistantship under Marston Morse at the Institute and an NYU instructorship with a four-year career in the U. S. Navy. He then was Associate Prof. and Prof. at the Univ. of N. Carolina before assuming his present position at Columbia. He spent leaves at Berkeley, Minnesota, Purdue, Michigan, and I.A.S.

He has served as President of the Institute of Math. Stat. and delivered the Rietz and Wald lectures. He was a Guggenheim Fellow in 1952–53.

Since his thesis in topology, Robbins's main research has been in probability theory and mathematical statistics. His book (with Y. S. Chow and D. O. Siegmund) *Great Expectations; The Theory of Optimal Stopping* will be published shortly by Houghton Mifflin. His previous *What is Mathematics?* with Richard Courant (Oxford Univ. Press, 1941) is a landmark in mathematical exposition. *Editor.*

Thus far we have said nothing about the nature of the reward sequence $x_n = f_n(y_1, \cdots, y_n)$. To begin with, let us take

$$(1) \qquad x_n = \frac{n2^n}{n+1} \cdot \prod_1^n \left(\frac{y_i + 1}{2}\right) \qquad (n = 1, 2, \cdots)$$

and analyze the resulting situation. Equation (1) is just a symbolic way of saying that if we stop after the nth toss with all heads we are to receive $n\,2^n/(n+1)$, while if any one of the first n tosses has been a tail we are to receive nothing. A little reflection will show that we need consider only the class of stopping rules $\{t_k\}$, $k=1, 2, \cdots$, where $t_k = k$; i.e., t_k stops after the kth toss no matter what sequence of heads and tails has appeared. Clearly

$$Ex_{t_k} = \frac{1}{2^k} \cdot \frac{k2^k}{k+1} + \left(1 - \frac{1}{2^k}\right) \cdot 0 = \frac{k}{k+1},$$

and therefore $V=1$ but no optimal stopping rule exists.

We remark in passing that at any stage n in which all heads have appeared, so that $x_n = n\,2^n/(n+1)$, the conditional expected reward for making one more toss before stopping is

$$E\left(x_{n+1} \middle| x_n = \frac{n2^n}{(n+1)}\right) = \frac{1}{2} \; \frac{(n+1)2^{n+1}}{(n+2)} = \frac{2^n(n+1)}{(n+2)} > x_n.$$

Hence it is always "foolish" to stop with all heads. But if we do not act "foolishly" at some point we shall wait for the first tail to occur and our final reward will always be 0. Thus acting "wisely" at each stage is the worst long-range policy.

EXAMPLE 2. The same, except that now

$$(2) \qquad x_n = \frac{y_1 + \cdots + y_n}{n} \qquad (n = 1, 2, \cdots).$$

This problem is much harder than the preceding one because of the enormous family of possible stopping rules which must be considered and evaluated. A simple instance of such a rule is

$$(3) \qquad t = \begin{cases} 1 & \text{if } y_1 = 1, \text{ otherwise} \\ n & \text{if } n \text{ is the first integer such that } y_1 + \cdots + y_n = 0. \end{cases}$$

The first question we must ask is, does (3) define a legitimate stopping rule in the sense that $P(t < \infty) = 1$? In probability theory it is shown that in repeated tossings of a fair coin, for any fixed integer $k = 0, \pm 1, \pm 2, \cdots$, the probability is 1 that the difference (number of heads in first n tosses) $-$ (number of tails in first n tosses) will assume the value k infinitely often. In our notation the case $k=0$ corresponds to the event $y_1 + \cdots + y_n = 0$, and hence $P(t < \infty) = 1$. It remains to evaluate Ex_t for (2) and (3). We have

$$Ex_t = \frac{1}{2}\cdot\frac{1}{1} + \frac{1}{2}\cdot 0 = \frac{1}{2}\cdot$$

It follows that $V \geqq \frac{1}{2}$, and of course $V \leqq 1$, since $x_n \leqq 1$ in all cases. By trial and error we can invent other stopping rules t for which $Ex_t > \frac{1}{2}$, but we will find none for which $Ex_t \geqq .9$, say. However, it is not so easy to *prove* that $V < .9$, still less to find the exact value of V and to determine whether an optimal t exists in this example.

In cases like this it is tempting to try to "put the problem on a computer." Briefly, here is what a computer can do. Suppose we restrict ourselves to the class C_N of all possible stopping rules t which take on only values in the set $\{1, 2, \cdots, N\}$, where N is some fixed positive integer; in other words, we restrict ourselves to stopping rules which always stop after at most N tosses of the coin. Denoting by v_N the supremum of Ex_t over the class C_N, we see that in this case C_N is a finite class and therefore an optimal rule in C_N must exist. Even so, for $N = 1000$, say, the class C_N is still too large for an exhaustive analysis. At this point the general theory of optimal stopping comes to our aid. Whenever the number of stages in the problem is bounded (even if the random variables involved have continuous rather than discrete distributions) there is always an optimal rule and a more or less constructive algorithm for finding it (cf. (22)). In the present problem a computer can actually be programmed to find v_N quite quickly for all N up to a few thousand. By definition the sequence v_N is nondecreasing, and therefore $v = \lim_{N\to\infty} v_N$ exists. But the values of v_N produced by a computer for successive values of N display no obvious pattern, and it is impossible even to guess the exact value of v from computer evidence, *still less to decide whether $V = v$ or $V > v$.*

In the present problem it has been proved [6, 11] that an optimal t does exist and that $v = V$, but the exact description of t and the value of V are not known.

EXAMPLE 3. As in Example 1 but now with

$$x_n = \min(1, y_1 + \cdots + y_n) - n/(n+1) \qquad (n \geqq 1).$$

Consider the stopping rule

$$t = \text{first } n \geqq 1 \text{ such that } y_1 + \cdots + y_n = 1. \tag{4}$$

That $P(t < \infty) = 1$ follows as in Example 2, and since $n/(n+1) < 1$,

$$Ex_t = 1 - E\left(\frac{t}{t+1}\right) > 0$$

(the exact value of Ex_t would require some probability theory to compute). A little thought will show that t is in fact optimal for this example and hence that $V = Ex_t > 0$.

On the other hand, since the y_i are independent and identically distributed

with $Ey_i=\frac{1}{2}\cdot 1+\frac{1}{2}\cdot(-1)=0$, Wald's lemma (see below) shows that if, unlike (4), t is any stopping rule for which $Et<\infty$, and hence in particular if $t\in C_N$ for some $N=1, 2, \cdots$, then $E(y_1+\cdots+y_t)=0$, so that

$$Ex_t \leqq E(y_1+\cdots+y_t) - E\left(\frac{t}{t+1}\right) \leqq -\frac{1}{2}\cdot$$

Hence, $v_N \leqq -\frac{1}{2}$ for all $N\geqq 1$, and hence $v=\lim_{N\to\infty} v_N \leqq -1/2$, while as we have seen $V>0$.

A slight variation on this example is given by

EXAMPLE 4. Let $y_1, y_2, \cdots$ be independent (but not identically distributed) random variables such that

$$P(y_i = 1 - a_i) = P(y_i = -1 - a_i) = 1/2, \quad \text{with } a_i = 1/i(i+1), \tag{5}$$

and let

$$x_n = y_1 + \cdots + y_n \qquad (n \geqq 1) \tag{6}$$

be the reward if we stop at the nth stage. Thus x_n represents the net gain of a gambler who plays a succession of unfavorable games,

$$Ey_i = \tfrac{1}{2}(1 - a_i) + \tfrac{1}{2}(-1 - a_i) = -a_i < 0,$$

and stops after the nth. It might seem that whatever stopping rule t he might use, his expected net gain after stopping, Ex_t, would be <0. However, let

$$t = \text{first } n \geqq 1 \text{ such that } \sum_1^n (y_i + a_i) = k, \tag{7}$$

where k is any preassigned positive integer. The argument in Example 2 again shows that $P(t<\infty)=1$, while by (6) and (7)

$$\begin{aligned} Ex_t &= E\left(\sum_1^t y_i\right) = k - E\left(\sum_1^t a_i\right) = k - E\left(\sum_1^t\left(\frac{1}{i} - \frac{1}{i+1}\right)\right) \\ &= k - E\left(\frac{t}{t+1}\right) > k - 1. \end{aligned}$$

Since k can be as large as we please, we see that $V=+\infty$ for this example. (The reader may decide whether there is an optimal t in this case, i.e., one for which $Ex_t=+\infty$.) We remark that although the one-step conditional expected reward $E(x_{n+1}|x_n)=x_n+E(y_{n+1})$ is always *less* than the present reward x_n, the use of a proper stopping rule makes the game profitable.

This example suggests the following question. If $y_1, y_2, \cdots$ are independent *and identically distributed* with $Ey_i<0$, does there exist a stopping rule t such that $E(\sum_1^t y_i)>0$? We shall now show that there does not.

We shall need the strong law of large numbers of probability theory.

THEOREM. (Kolmogorov) *Let $y_1, y_2, \cdots$ be independent and identically distributed random variables with the common distribution function $F(y)=P(y_i<y)$, and put $x_n=\sum_1^n y_i$. If*

$$Ey_i = \int_0^\infty ydF(y) + \int_{-\infty}^0 ydF(y) \tag{8}$$

exists (i.e., if the two integrals in (8) *are not respectively $+\infty$ and $-\infty$), denote it by μ. Then*

(a) *If μ exists, $-\infty \leq \mu \leq \infty$, then $P(\lim_{n\to\infty}(x_n/n)=\mu)=1$.*

(b) *If for some finite constant c,*

$$P\left(\lim_{n\to\infty}\frac{x_n}{n} = c\right) = 1, \tag{9}$$

then μ exists and equals c.

We remark parenthetically that (9) may hold with $c=+\infty$ or $-\infty$ even though μ does not exist.

Now let t be any stopping rule for an independent and identically distributed sequence $y_1, y_2, \cdots$ for which $\mu=Ey_i$ exists, $-\infty\leq\mu\leq\infty$, and consider the randomly stopped sum $x_t=\sum_1^t y_i$. For any $n\geq 1$ let

$$S_n = (y_1 + \cdots + y_{t_1}) + (y_{t_1+1} + \cdots + y_{t_1+t_2}) + \cdots + (y_{t_1+\cdots+t_{n-1}+1} + \cdots + y_{t_1+\cdots+t_n}),$$

where $t_1=t$, t_2 is t applied to the sequence $y_{t_1+1}, y_{t_1+2}, \cdots$, etc. It is easy to see that the n groups of terms in S_n are in fact n independent random variables with the same probability distribution as x_t, and that $t_1, t_2, \cdots$ are independent random variables with the same distribution as t. Since $\mu=Ey_i$ exists by hypothesis and since Et always exists (since $t\geq 1$), it follows from (a) that with probability 1 as $n\to\infty$

$$\frac{S_n}{n} = \left\{\frac{y_1 + \cdots + y_{t_1+\cdots+t_n}}{t_1+\cdots+t_n}\right\}\cdot\left\{\frac{t_1+\cdots+t_n}{n}\right\} \to \mu\cdot Et \tag{10}$$

provided only that the last expression is not of the form $0\cdot\infty$. Applying (a) again to $x_t=\sum_1^t y_i$ we have [15] the

COROLLARY. *If $y_1, y_2, \cdots$ are independent and identically distributed, if $\mu=Ey_i$ exists, and if t is any stopping time of the sequence $y_1, y_2, \cdots$ for which $E(\sum_1^t y_i)$ exists, then*

$$E\left(\sum_1^t y_i\right) = \mu\cdot Et \tag{11}$$

whenever the right side of (11) *is not of the form $0\cdot\infty$.*

From (b) and (10) we obtain

WALD'S LEMMA [18]. *If* μ *and* Et *are both finite then* $E(\sum_1^t y_i)$ *always exists and satisfies* (11).

Wald's Lemma admits of many generalizations. For example, if $y_1, y_2, \cdots$ are independent but not necessarily identically distributed random variables for which $E|y_i| \leqq C < \infty$ for all $i \geqq 1$, if $\mu_i = Ey_i$, and if t is any stopping rule of the sequence $y_1, y_2, \cdots$ such that $Et < \infty$, then $E(\sum_1^t y_i)$ exists and $E(\sum_1^t y_i) = E(\sum_1^t \mu_i)$. Applied to Example 4 in which $|y_i| \leqq 2$ we see that if in that example t is any stopping time of the sequence $y_1, y_2, \cdots$ such that $Et < \infty$, then

$$Ex_t = E\left(\sum_1^t \mu_i\right) = -E\left(\sum_1^t a_i\right) = -E\left(\frac{t}{t+1}\right) \leqq -\frac{1}{2}\cdot$$

Thus to achieve $Ex_t > -1/2$ as we did with (7), we have to use a t for which $Et = \infty$. This detracts somewhat from the prospect of getting rich in the long run by using (7) or something like it.

We are now able to justify our negative answer to the question raised in the paragraph following Example 4. The Corollary shows that if $y_1, y_2, \cdots$ are independent and identically distributed with $Ey_i = \mu < 0$, and if t is any stopping rule for which $E(\sum_1^t y_i)$ exists (even though Et may be $+\infty$), then

$$E\left(\sum_1^t y_i\right) = \mu\cdot Et < 0; \tag{12}$$

the optimal stopping rule is therefore $t \equiv 1$, and $V = \mu < 0$. (We remark that there may exist t's such that $E(\sum_1^t y_i)$ does not exist, but such t's are excluded from consideration under our definition of optimality.)

Our next example is of a different character from the preceding ones in that it can be supposed to apply to certain problems in real life rather than to imaginary gambling situations.

EXAMPLE 5. Let $y_1, y_2, \cdots$ be independent random variables with a common distribution function $F(y) = P(y_i < y)$, and let

$$x_n = \max(y_1, \cdots, y_n) - cn \qquad (n \geqq 1) \tag{13}$$

where c is some positive constant. What stopping rule t maximizes Ex_t? A complete solution to this question is given by

(i) If $\int_0^\infty y\,dF(y) = \infty$ then an optimal stopping rule is

$$t = \text{first } n \geqq 1 \text{ such that } y_n \geqq b \ (b \text{ any finite constant}), \tag{14}$$

and $V = Ex_t = +\infty$, while

(ii) if $\int_0^\infty y\,dF(y) < \infty$ there exists a unique number β such that

$$\int_\beta^\infty (y - \beta)\,dF(y) = c; \tag{15}$$

an optimal stopping rule is

$$t = \text{first } n \geqq 1 \text{ such that } y_n \geqq \beta, \tag{16}$$

and $V = Ex_t = \beta$.

Proof. (i) Let $p = p(y_i \geqq b) = 1 - F(b) > 0$ and $q = 1 - p$. Then for (14)

$$P(t = n) = pq^{n-1}, \qquad P(t < \infty) = \sum_1^\infty pq^{n-1} = p \cdot \frac{1}{1-q} = 1,$$

$$Et = \sum_1^\infty npq^{n-1} = p\frac{d}{dq}\left(\sum_0^\infty q^n\right) = p \cdot \frac{1}{(1-q)^2} = \frac{1}{p}, \tag{17}$$

and

$$E \max(y_1, \cdots, y_t) = \frac{\int_b^\infty ydF(y)}{p} = \infty.$$

Hence by (17)

$$Ex_t = \infty - \frac{c}{p} = \infty,$$

so t is necessarily optimal.

(ii) Consider the function

$$\phi(b) = \int_b^\infty (y - b)dF(y),$$

which is equal to the area of the region in a y, z-plane bounded below by the curve $z = F(y)$, on the left by the line $y = b$, and above by the line $z = 1$. It is geometrically evident that there is a unique solution of (15) for any given $c > 0$, and for t defined by (16) we have as in (i)

$$Et = \frac{1}{p}, \qquad Ex_t = \left(\int_\beta^\infty ydF(y) - c\right)/p \qquad (p = 1 - F(\beta)).$$

Hence by (15)

$$Ex_t = \beta. \tag{18}$$

To prove that the t defined by (16) is optimal, let t' be any stopping rule such that $Ex_{t'}$ exists and is $> -\infty$. Choose any $b > \beta$ and observe that

$$x_n = \max(y_1, \cdots, y_n) - cn \leqq b + \sum_1^n ((y_i - b)^+ - c), \tag{19}$$

where we define $a^+ = \max(a, 0)$. The sequence of random variables $w_1, w_2, \cdots$, where $w_i = (y_i - b)^+ - c$, is independent and identically distributed with $\mu = Ew_i = \int_{-\infty}^{\infty} ((y-b)^+ - c)dF(y) = \phi(b) - c < 0$ since $b > \beta$ and β satisfies (15). Now $Ex_{t'} > -\infty$ by hypothesis, so by (19)

$$E\left(\sum_1^{t'} w_i \right) \geqq Ex_{t'} - b > -\infty, \tag{20}$$

and hence by Corollary 1, whether Et' be finite or infinite,

$$E\left(\sum_1^{t'} w_i \right) = \mu Et' < 0; \tag{21}$$

thus by (20) and (21) $Ex_{t'} < b$. Since b was any constant $> \beta$ this implies that $Ex_{t'} \leqq \beta$ and hence by (18) that t is optimal and $V = \beta$.

The original proof of this result [4] along lines suggested by the general theory was considerably more complicated and was valid only under the unnecessary hypothesis that $\int_0^\infty y^2 dF(y) < \infty$. The direct proof just given is based on the trivially simple inequality (19) which reduces the problem from one involving the maximum of $y_1, \cdots, y_n$ to one involving a simple sum. This inequality was pointed out to the author by David Burdick.

We remark that the optimal stopping rule (16) in case (ii) is suggested by the one-step argument which would have led us astray in Examples 1 and 4 (and in Example 3, also for that matter). For, putting $m_n = \max(y_1, \cdots, y_n)$ we have

$$\begin{aligned} x_{n+1} &= m_{n+1} - c(n+1) = \max(m_n, y_{n+1}) - c(n+1) \\ &= m_n + (y_{n+1} - m_n)^+ - c(n+1) \\ &= x_n + (y_{n+1} - m_n)^+ - c, \end{aligned}$$

so that $E(x_{n+1} | y_1, \cdots, y_n) = x_n + E(y_{n+1} - m_n)^+ - c > x_n$ if and only if

$$\int_{m_n}^{\infty} (y - m_n) dF(y) > c;$$

i.e., if and only if $m_n < \beta$.

We remark also that if we replace (13) by $\bar{x}_n = y_n - cn$ $(n \geqq 1)$, the optimal solution remains the same, since

$$\bar{x}_n \leqq x_n \quad \text{but} \quad \bar{x}_t = x_t$$

with t defined by (14) or (16).

The general theory of optimal stopping is more than a collection of examples and *ad hoc* methods of solution. It draws heavily on martingale theory and the other apparatus of modern probability theory, to which it also contributes. Rather than sketch the general methods in a necessarily incomplete form we have preferred to present a few simple examples in the hope that the reader will

be induced to consult some of the references below for a more formal presentation.

Our last example will be one in which the number of random variables is finite from the outset and hence no difficulties of principle arise concerning the existence of an optimal solution, but the explicit result is none the less surprising.

EXAMPLE 6. An employer interviews a finite number N of applicants for a position. They are Miss 1, Miss 2, $\cdots$, Miss N, in decreasing order of excellence, and they are interviewed in random order, each of the $N!$ permutations being equally likely. The rules of the game are as follows. After each interview the employer can rank the girl just interviewed relative to her predecessors but is ignorant of her absolute rank relative to the whole group of N. Thus if the first few girls in order of appearance were Misses

$$7, 3, 9, 4, \cdots$$

the employer would know successively that

the first girl has relative rank 1
the second has relative rank 1
the third has relative rank 3
the fourth has relative rank 2, etc.

The employer must hire one applicant, but if at any stage he does not hire the girl just interviewed he must dismiss her and cannot call her back later. *The object of the employer is to minimize the expected absolute rank of the girl hired.*

If t is any stopping rule for this problem and x_t denotes the absolute rank of the girl hired using t, then the possible values of x_t range from 1 (best girl hired) to N (worst hired). If t consists of always hiring the first girl to appear (or hiring at random one of the N girls) then

$$Ex_t = (N + 1)/2,$$

since the first (or a randomly chosen) applicant is equally likely to have any of the absolute ranks 1, 2, $\cdots$, N. Clearly, a good strategy t, based on the information on relative ranks, should be able to do better than that. Denote

$$R_N = \inf_t\{Ex_t\} \qquad (N \geqq 1),$$

where the inf is taken over all the (finite) number of possible strategies available with a group of size N. For example, when $N=1$ only one strategy is available and $R_N=1$, for $N=2$ there are only two possible strategies, equally good, and $R_2=\frac{3}{2}$, for $N=3$ there are four strategies available of which the best gives $R_3=\frac{5}{3}$, etc.

It is possible by using the general theory to devise a computer program for finding R_N for any given $N=1, 2, \cdots$. It amounts to the following: Define

$$c_{N-1} = (N + 1)/2, \quad \text{and for } i = N - 1,\ N - 2, \cdots, 1$$

compute successively

$$(22)\qquad \begin{cases} s_i = \left[\dfrac{i+1}{N+1}c_i\right], \quad \text{where } [x] = \text{largest integer} \leqq x, \quad \text{and} \\ c_{i-1} = \dfrac{1}{1}\left(\dfrac{N+1}{i+1}\cdot\dfrac{s_i(s_i+1)}{2} + (i-s_i)c_i\right). \end{cases}$$

Then $R_N = c_0$. The backward induction from c_{N-1} to c_0 is an example of the general algorithm for solving optimal stopping problems when there are only a finite number of successive stages $1, \cdots, N$ at which stopping is permitted. In the present problem, c_i represents the minimal expected absolute rank attainable by any stopping rule which must stop somewhere between the $(i+1)$th and Nth stages inclusive.

Using this procedure it is found that

$$(23)\qquad R_{10} = 2.56, \quad R_{100} = 3.60, \quad R_{1000} = 3.83.$$

(The optimal strategy in each case is also produced by the computer but we shall not bother to describe it here.) No general formula for R_N other than the recursive definition by way of (22) is available, and for each N the recursion has to be started afresh.

Inspection of (23) shows that the values R_N seem to be increasing, but surprisingly slowly. It can in fact be proved that R_N increases steadily with N and hence that the constant

$$R = \lim_{N\to\infty} R_N \leqq +\infty$$

is well defined. What is its value, finite or infinite? In [3] it is proved that

$$R = \prod_{j=1}^{\infty}\left(\frac{j+2}{j}\right)^{1/(j+1)} \cong 3.8695.$$

Thus, no matter how large N is, there exists a strategy for which the expected absolute rank of the girl hired is less than 4, even though the actual absolute rank of the girl hired may be as great as N.

If instead of arriving in random order with each permutation equally likely, the applicants are sent in by a malicious competitor of the employer who wants to force him to do badly, the competitor can use certain permutations more often than others in such a way that no matter what strategy the employer uses (even knowing the competitor's mode of randomization) he will always have expected absolute rank of applicant hired $=(N+1)/2$, the purely random expectation. This is relevant to certain real or imaginary military applications of the hiring problem.

There are many variants [12] of this problem, the most common being that in which the employer, instead of trying to minimize the expected absolute rank

of the girl hired, tries to maximize the probability of hiring the best girl, regardless of whom he hires when he does not get the best. For this case it has long been known that a probability $\cong 1/e$ of hiring the best can be attained by the proper strategy. The problem as we have put it is, of course, more difficult and perhaps more interesting.

References

1. K. J. Arrow, D. Blackwell, and M. A. Girshick, Bayes and minimax solutions of sequential decision problems, Econometrica, 17 (1949) 213–244.

2. L. Breiman, Stopping-rule problems, Applied Combinatorial Mathematics (ed. E. F. Beckenbach), Wiley, New York, 1964, pp. 284–319.

3. Y. S. Chow, S. Moriguti, H. Robbins, and S. M. Samuels, Optimal selection based on relative rank, Israel J. Math., 2 (1964) 81–90. The problem was stated by A. Cayley, "Mathematical questions and their solutions," Educational Times, 22 (1875) 18–19.

4. Y. S. Chow and H. Robbins, A martingale system theorem and applications, Proc. Fourth Berkeley Symp. on Math., Stat. and Prob., 1 (1960) 93–104.

5. ——— and ———, On optimal stopping rules, Zeitsch. f. Wahr., 2 (1963) 33–49.

6. ——— and ———, On optimal stopping rules for s_n/n, Illinois J. Math., 9 (1965) 444–454.

7. ——— and ———, On values associated with a stochastic sequence, Proc. Fifth Berkeley Symp. on Math. Stat. and Prob., 1 (1965) 427–440. A class of optimal stopping problems, *ibid.*, 419–426.

8. ———, H. Robbins, and D. O. Siegmund, Great Expectations: The Theory of Optimal Stopping, Houghton Mifflin, Boston, to appear in 1970.

9. ———, H. Robbins, and H. Teicher, Moments of randomly stopped sums, Ann. Math. Statist., 36 (1965) 789–799.

10. J. L. Doob. Stochastic Processes, Wiley, New York, 1953.

11. A. Dvoretzky, Existence and properties of certain optimal stopping rules, Proc. Fifth Berkeley Symp. on Math. Stat. and Prob., 1 (1965).

12. J. P. Gilbert and F. Mosteller, Recognizing the maximum of a sequence, J. Amer. Statist. Ass., 61 (1966) 35–73.

13. G. W. Haggstrom, Optimal stopping and experimental design, Ann. Math. Statist., 37 (1966) 7–29.

14. ———, Optimal sequential procedures when more than one stop is required, *ibid.*, 38 (1967) 1618–1626.

15. H. Robbins and E. Samuel, An extension of a lemma of Wald, J. Appl. Prob., 3 (1966) 272–273.

16. D. O. Siegmund, Some problems in the theory of optimal stopping rules, Ann. Math. Statist., 38 (1967) 1627–1640.

17. J. L. Snell, Application of martingale system theorems, Trans. Amer. Math. Soc., 73 (1952) 293–312.

18. A. Wald, On cumulative sums of random variables, Ann. Math. Statist., 15 (1944) 283–296.

REMINISCENCES OF A MATHEMATICAL IMMIGRANT IN THE UNITED STATES

SOLOMON LEFSCHETZ, Princeton University and Brown University

My career as mathematical immigrant began in 1911 upon my receiving the Ph.D. degree from Clark University (Worcester, Mass.). While small, Clark had as its President G. Stanley Hall, an outstanding psychologist, and several distinguished professors. The mathematical faculty consisted of three members: W. E. Story, senior professor (higher plane curves, invariant theory); Henry Taber (complex analysis, hypercomplex number systems); De Perrott (number theory).

There were great advantages for me at Clark. I graduated from the *École Centrale* (Paris) (one of the French "*Grandes Écoles*") in 1905, and for six years was an engineer. I soon realized that my true path was not engineering but mathematics. At the *École Centrale* there were two Professors of Mathematics: Émile Picard and Paul Appel, both world authorities. Each had written a three-volume treatise: *Analysis* (Picard) and *Analytical Mechanics* (Appel). I plunged into these and gave myself a self-taught graduate course. What with a strong French training in the equivalent of an undergraduate course, I was all set.

To return to Clark, I soon obtained a research topic from Professor Story: to find information about the largest number of cusps that a plane curve of given degree may possess. An original contribution which I made secured my Ph.D. thesis and my doctorate in 1911.

At Clark there was fortunately a first rate librarian, Dr. L. N. Wilson, and a well-kept mathematical library. Just two of us enjoyed it—my fellow graduate student in mathematics and future wife, and myself. I took advantage of the library to learn about a number of highly interesting new fields, notably about the superb Italian school of algebraic geometry.

Prof. Lefschetz continues an astonishingly productive career. His profound influence in the development of topology and of algebraic geometry is expounded at length in articles by W. V. D. Hodge and Norman E. Steenrod in the Princeton Symposium volume in honor of S. Lefschetz, *Algebraic Geometry and Topology* (1957) edited by R. H. Fox, D. C. Spencer, and A. W. Tucker. His numerous publications in these fields include the books, *L'Analyse Situs et la Géométrie Algébrique* (1924), *Géométrie sur les Surfaces et les Variétés Algébriques* (1929), *Topology* (1930), *Algebraic Topology* (1942), *Topics in Topology* (1942), *Introduction to Topology* (1949), and *Algebraic Geometry* (1953). In recent years he has produced fundamental research in ordinary differential equations, including the volumes *Differential Equations, Geometric Theory* (1957), and (with J. La Salle) *Stability by Liapunov's Direct Method with Applications* (1961).

Prof. Lefschetz began his mathematical career in 1911 with his PhD under W. E. Story at Clark University. He held positions at the Univ. of Nebraska, Univ. of Kansas, then Princeton University until his retirement. At Princeton he was Research Professor, 1932–1953, and Department Chairman, 1945–1953. Since, he has been at the National University of Mexico, RIAS, and Brown University. His numerous awards include the Bordin Prize (Académie des Sciences 1919), the Bôcher Prize (AMS 1924), the Feltrinelli Prize (Academia dei Lincei 1956), and foreign memberships in the Royal Society and the Académie des Sciences. He was Editor, Annals of Mathematics, President AMS (1935–1937), and is a member of the National Academy of Sciences and the American Philosophical Society. *Editor.*

My first position was an assistantship at the University of Nebraska (Lincoln), soon transformed into a regular instructorship. This meant my first contact with a regular midwestern American institution and I enjoyed it to the full. I owed it mainly to the very pleasant and attractive head of the department, Dean Davis of the College. The teaching load, while heavy, did not overwhelm me since it was confined to freshman and sophomore work.

Not too many weeks after my arrival, the Dean got me to speak before a group of teachers in Omaha on "Solutions of algebraic equations of higher degree." And then and there I learned an all-important lesson. For I spoke three quarters of an hour—three times my allotted time! When I found this out some weeks later from the Dean, my horror knew no bound. I decided "never again," to which I have most strictly adhered ever since.

A second lesson was of another nature. I utilized my considerable spare time in reading Hilbert's recent papers on integral equations. At Clark I had also read Fredholm's Acta paper on the same topic and my enthusiasm for integral equations was very great. I offered to lecture on Hilbert's work in my fourth term and this was accepted. Consequence: a very heavy teaching load for two students who I fear were quite bewildered. One of them, Oliver Gish, a graduate student in physics (later a distinguished geophysicist) remained my lifelong friend. I also formed a close friendship with his mentor and a capable mathematician, Professor L. B. Tuckerman (later of the Bureau of Standards).

The course taught me a valuable lesson: the experience generally absorbs too much energy. I have since expressed this opinion to many a recent doctor, but I fear that few heeded it.

My two years in Nebraska made me realize a widespread feature of American institutions of higher learning which were State institutions. By general state rule they had to accept any graduate from an accredited high school. Consequence: in the freshman year a flood of very poorly prepared students and a large number of sections, especially in the first term. By the end of the first year the entrance flood was reduced to half; the sophomore sections—in mathematics at least—were in much smaller number, more readily handled and better taught. This went on down to the last year, with the flood in mathematics reduced to 10–15 or so (mostly girls) and the total number of graduates much smaller than at entrance.

Lincoln, the capital of the State (population about 50,000) was a very pleasant city, with a distinct urban flavor. It was not too far from Omaha, the major city of the State. Most family houses were surrounded by a small garden and the whole made a very good impression. The University was at one end of town; the Agricultural College, part of the University (pet of the very rural Board of Regents), at the other end. There were a couple of small colleges situated in Lincoln.

At the end of two years (1913) a larger offer, plus my approaching marriage to my Clark fellow student, made me accept an instructorship at the University of Kansas in Lawrence. The teaching conditions there were the same as in

Lincoln, but with a slightly smaller load. At the University of Kansas the department was divided into two groups: college plus graduate work and engineering. I was assigned to the latter. While the students were somewhat more purposeful, the preparation was equally weak in both parts.

Lawrence (population 12,000) had a rather severe New England tradition. Except for the University with about 3000 students, it was really a most pleasant rural community. The University was on top of quite a hill, with well-constructed and mostly recent buildings. The view from the top was exceptionally attractive.

The major city near by was Kansas City. Lawrence was about 25 miles from Topeka (the capital), while Kansas City was 50 miles away. This was all before the automobile age and my friends and I indulged in many country walks.

The general entrance preparation in Lawrence and Lincoln was so feeble that early teaching could only be technical and deprived of theory. As the freshman flood eroded, this situation improved somewhat.

The rule in Lawrence for beginning faculty members was three years in each position and it was rather rigidly enforced. The situation did not seem perfect—far from it. However, I discovered in myself, first a total lack of desire to "reform" coupled with a large adaptive capacity. At Lawrence I only cooperated with a colleague in driving out several unattractive texts, notably Granville's Calculus, for which my taste was $<\epsilon$.

Years later I inquired of Professor Lusin (Moscow) why the Soviet mathematicians translated Granville. Reply: "We only took his excellent collection of problems, but provided our own theory." This may explain our efforts to move this book out of Kansas.

At this place I was prepared to indulge in extensive criticism, at least of the midwestern system. The fact is, however, that in both Nebraska and Kansas I found good and well-kept mathematical libraries, ample at least for my own purposes. Moreover, I came to realize the enormous advantage over the European system: it provided uncountably many opportunities for younger research men with ideas to grow and develop their powers, as instructors for example, with ample leisure. For the teaching loads, while considerable, were not really intolerable. Moreover, they generally went with colleagues who had other interests, mathematical or administrative, but not intent upon imposing on one uncongenial mathematical interests. At all events, in my case, it turned out to be of great value. Needless to say, special research favors were rare indeed.

In spite of the general level, I had in Lawrence three or four excellent students. One of them, Warren Mason, went to work for Bell Laboratories in New York (later near Elizabeth, N. J.), took his Ph.D. in physics at Columbia, and at Bell became a top specialist in the theory of sound and its applications. I am very proud of him. Still another strong student, Clarence Lynn, joined forces with Westinghouse in Pittsburgh (electrical department) and was most successful there.

I have found that in freshman courses in mathematics, and less so in the

next year, hardly one third of the students care for and are not totally bored by mathematics. Hence at that early level a teacher must be exceptionally lively and have a sympathetic understanding of the students. Needless to say this must be coupled with a complete grasp of the topic taught.

Here are a few very radical suggestions for later years. From the junior year on through graduate work they should be merged into a professional school, with teaching, at least in mathematics, of seminar type plus abundant but easy contact with faculty on an individual basis. In other words "baby talk" should end with sophomore years.

The guidelines in my research were: Picard-Simart: *Fonctions algébriques de deux variables* (two volumes, mostly Picard); Poincaré's papers on topology (=analysis situs) and on algebraic surfaces; Severi's two papers on the theory of the base; Scorza's major paper (dated 1915) in *Circole di Palermo* on Riemann matrices.

Around 1915 and for a long time, a certain result of Picard baffled me. Let H be a hyperelliptic surface. Direct calculation yielded: the Betti number $R_2(H)=6$. Picard, however, appeared to give its value as 5. The discovery of the missing link played a major role for me. Namely, Picard only wanted R_2 for the *finite* part of H, neglecting the curve C *at infinity. Hence* C *was a* 2-*cycle*, and so was any algebraic curve! This launched me into Poincaré-type topology, the 1919 Bordin Prize of the Paris Academy and in 1924 Princeton! (The translated prize paper appeared in the Trans. Amer. Math. Soc., vol. 22, 1921.)

The immediate effect of the Prize was the Kansas promotion (January 1920) to Associate Professor plus a schedule reduction. Also (1923) there came a promotion to a Full Professorship. I spent the year 1920–21 in Europe, half in Paris, half in Rome. I gathered little mathematical profit in Europe; some from the summer of 1921 which I spent in Chicago.

About Paris I particularly remember an interview with Émile Borel lasting five minutes in which I offered to write for his Series my future monograph *L'Analysis Situs et la Géométrie Algébrique*. He accepted at once! (In such matters our "speedy" country knew no such speed.) Proof sheets, etc., were dealt with rapidly and not a syllable was changed.

I come now to my Princeton period. In 1923 an invitation came from Dean Fine, the Chairman of the Department of Mathematics and Dean of the Faculty at Princeton, to spend the following year there as Visiting Professor of Mathematics. Dean Fine was the long-time head of the department and the true founder of what became an outstanding department of the University. With reason, upon the construction of the mathematical building it was called "Fine Hall." (Dean Fine was killed in an automobile accident just before Christmas 1928 and his lifelong friend, Mr. Thomas D. Jones, immediately granted $600,000 as a memorial to Dean Fine for a new mathematical building.)

Well, upon receiving Dean Fine's invitation, I accepted. For the following year I received a permanent offer to stay at Princeton as Associate Professor.

This was changed 18 months later (January 1927) to a Full Professorship and January 1932 to a Research Professorship (Fine professorship) as successor to Oswald Veblen. In this position I had no assigned duties whatever.

At Princeton I found myself in a world-renowned University and in one of its outstanding Departments. Among the great mathematical Professors there were: Eisenhart, Veblen, Wedderburn, Alexander, Hille. I was in closest contact with Alexander—a top authority in topology.

My joining the Princeton faculty coincided with a definite change of direction in my research from the applications of nascent topology to algebraic geometry (*vide* my prize paper) to a pure topological problem: coincidences and fixed points of transformations. For this problem I invented a completely new method of attack, which by 1925 culminated in a well-known fixed point expression $\phi(f)$, f a mapping of a manifold into itself, that said: $\phi \neq 0$ implies that f has fixed points; if f has none, $\phi = 0$. The preparation and extensions required occupied me for several years. One of my early graduate students, A. W. Tucker, an outstanding Princeton mathematician, found the way to a far simpler method than my early one, which I have accepted *in toto*.

Much of my Princeton teaching, until 1930, was still freshman-sophomore. However the students, selected with care at entrance, were much better prepared than in the midwest. The contrast of the systems was very great.

Princeton system: A strictly private school, with limited funds and space, could not accept all comers. Hence it had, unavoidably, to fix the number of admissions, utilize a strict selection, and keep the admitted men practically through the four collegiate years. The same system, in some form, was also applied to admission to the Graduate School.

Midwestern system: As I already stated, they had to admit all duly certified high school students. The freshman entrance flood resulted in teaching mostly by graduate students, many of uncertain quality.

The Princeton system had two important consequences. First, it enabled one to organize preferred sections even before entrance. Second, courses could be initiated at a more advanced stage and proceed more speedily. Thus algebra and trigonometry were done each in two weeks, analytical geometry in five weeks, calculus started in the second freshman semester (in Kansas-Nebraska in the sophomore year).

Some years later, good students from strong preparatory schools or high grade secondary schools (where they already had these subjects) were allowed to skip, even the whole first year. Moreover, such A-1 men (not many) were soon treated like graduate students, allowed to participate in advanced seminars and thus to become well acquainted with the members of the mathematical faculty.

The Princeton aim was decidedly different from the Nebraska-Kansas aim. The latter had to provide for a considerable number of teachers in their states, to form moderate level technicians of all kinds, sending a very few of the best

for better training to major eastern institutions. Princeton on the contrary was planned to form the top echelons, notably in the sciences. This meant aiming first for the doctorate. In mathematics it soon became customary to retain the best men for at least one year after the Ph.D. on some fellowship, or in some teaching position with very light duties. A number of the men so developed occupy today major posts in outstanding institutions.

In 1932 a major change took place through the establishment at Princeton of the Institute for Advanced Study, with mathematics as its first and strongest group. This resulted in the migration of three of our major members: Veblen, Alexander and von Neumann.

The basic effect on me was regaining the mathematical calm of Nebraska-Kansas, which I had so enjoyed without realizing it. Our mathematics chairman, Dean L. P. Eisenhart, with the unstated motto "live and let live" had much to do with this return of calm. During this period my mathematical work progressed. My first Topology treatise (1930) appeared and was many times approved by friendly colleagues. A second Algebraic Topology appeared in 1942, rather less satisfactory, because too algebraic. Other books came. I was editor of the *Annals of Mathematics*, which grew to occupy an A-1 place in mathematics, but did not overwhelm me with work. Then came World War II and I turned my attention to Differential Equations. With Office of Naval Research backing (1946–1955) I conducted a seminar on the subject from which there emanated a number of really capable fellows, also a book: *Differential Equations, Geometric Theory* (1957).

When Dean Eisenhart retired (1945) I succeeded him as Chairman, until my own retirement in 1953.

In 1944 I joined as a part-time connection the *Instituto de Mathematicas* at the National University of Mexico. This continued until 1966. At the *Instituto* I was as free as under my Princeton professorship. I conducted seminars in topology and differential equations, gave a couple of times a "volunteer" course on "general mathematical concepts" directed at beginners and, thanks to a good working library, was able to continue research. Conditions were of course quite different from ours, but as I became rapidly fluent in Spanish, it gave me many advantages. Through the years I found quite a number of capable young men, several of whom I directed to Princeton for further advanced training up to the doctorate and later. Among them I may mention Dr. José Adem, Chairman of the Department of Mathematics of the newly founded *Centro de Estudios Avanzados* in Mexico City.

My long connection with Mexico has been the occasion of many side trips (especially in connection with meetings of the Mexican Mathematical Society), so that I have a fair acquaintance with that wonderful country.

In 1964 the rarely awarded order of the Aztec Eagle was conferred upon me by the government of Mexico.

My work as Russian reviewer for differential equations had made me aware

of our lag relative to the Soviets in this all important field in all sorts of applications. The arrival of Sputnik in 1957 convinced me that this lag had to be remedied. As I attributed it to our scattered efforts, I came to the conviction that the only remedy was to establish a Center for study and research in differential equations.

From Dr. Robert Bass, formerly a member of my project, I learned of the formation in Baltimore, as a division of the Martin Aircraft Company, of a new Research Institute for Advanced Study (RIAS) under the direction of Welcome Bender, a graduate of MIT and long time Martin engineer. When I approached him with my (modest) plans he was enthusiastic. In a few days I was entrusted with the formation of a group of say five top men and about ten younger associates, with myself as director. Suffice it to say that I had considerable success. I first was able to obtain the cooperation of Prof. Lamberto Cesari of Purdue, one of the major specialists anywhere; also of Notre Dame, Prof. J. P. Lasalle as my second in command (my best appointment) and complete the group with Dr. J. K. Hale of Purdue (Cesari's best student there) and Dr. Rudolph Kalman of Columbia (an electrical engineer coupled with good mathematics). My strong basic group was thus complete.

I demanded (and obtained) from Mr. Bender that my group operate under standard university conditions.

Very shortly we became known. A considerable number of the good differential-equationists visited us, and some few were invited for a year or so.

After some six years it was necessary to transfer our Center elsewhere. This operation, carried by Lasalle, resulted in our becoming part of the Division of Applied Mathematics at Brown University as "Center for dynamical systems" with Lasalle as Director and myself as (once weekly) Visiting Professor. At Brown our general relationship has been excellent. A year or so ago the Director of the Division died and was succeeded by Lasalle whose general performance could not be excelled.

In conclusion I must recognize a budget of debts which I may never succeed in liquidating to the full.

The first is my enormous debt to my wife Alice, my Clark companion. Without her constant and unfailing encouragement through 59 years, 56 as my wife, I would have long since ceased to operate. . . .

Second major debt: to the United States, which through their (however imperfectly organized) universities made it possible for me to follow my deep bent for mathematics. I should also include here the contribution of the National University of Mexico from 1944 to 1965—years after my Princeton retirement, and also of RIAS and Brown.

In this long and agreeable route of 57 years I encountered so many *simpáticos amigos* that to name them all would be impossible. May they one and all accept my fervent *gracias* for my debt to them. I hope that they have felt that it was not incurred in vain.

STUDENT ORIENTED TEACHING—
THE MOORE METHOD

LUCILLE S. WHYBURN, University of Virginia

"I just never took Point Set Topology until I graduated and took a Summer course from Moore, quite by accident, and liked him. I liked the excitement in his class, the fact that we were working. It was different from other classes. I started to understand fine details about things, which attracted me." This student view of R. L. Moore's classes contrasts sharply with Robert S. Morison's statement before the AAAS-BAAS Conference in Boulder, Colorado, during April 1969: "My over-all impression, in returning to a university after a lapse of twenty years, is one of disappointment that so few students seem to have much fun either in their science courses specifically or in university life in general." The loyalty of his current students affords Professor Moore the opportunity to continue teaching in 1969 at the University of Texas though he is well past the usual age of mandatory retirement. He carries a fifteen hour teaching load of five courses. His offering consists of a course in Calculus and four courses chosen from the following listing:

Introduction to the Foundations of Analysis;
Introduction to the Foundations of Geometry;
Theory of sets;
Foundations of Mathematics;
Point Sets and Continuous Transformations;
Research in Point Set Theory.

As one of the prerequisites the catalog lists "consent of the instructor," a prerequisite that is not to be taken lightly. "Consent of the instructor" implies that the student has attained a certain stage of mathematical maturity, that he is interested in "doing" mathematics not in ferreting out what former mathematicians have done through the reading of theorems and proofs or the application of knowledge thus gained to problems. Passive listening followed by exhibition on quizzes and examinations that he has understood what was said is not sufficient in this class, each member must want a piece of the action. The discovery of proofs and definition of concepts unknown to the student under strongly competitive stimuli challenge the student to exercise his own honor system about not reading related material or cribbing ideas from any source.

In the popular mind the MOORE-METHOD is the axiomatic method. Certainly this is a part of the method and it is a portion that is easily accepted by and has wide appeal for mathematicians. The proliferation of books in general topology, modern algebra, and set theory address themselves to the axiomatic method. Many of them are excellent. Such books have presented new insights into

mathematical relationships, caused re-examination of curriculums, and introduced a new spirit of adventure into the teaching of mathematics. The essence of the Moore-Method, however, is to use an axiomatic treatment to create in the student a spirit of self-confidence and pleasure in personal creative endeavor. Much depends upon the handling of the material. The present widespread effort to use axiomatic systems at many levels of mathematical instruction has met with spotty success. From Professor Moore's point of view, his first fruitful contact with a student is in the calculus. He is a thorough teacher of manipulative skills commonly emphasized in such courses but in addition he is likely to introduce some use of syllogisms and a hard look at Duhamel's Theorem in order to afford the better students an opportunity to reason carefully. This challenge is not aimed at, nor does it attract, the indifferent or pedantic student.

No textbook, no description, can recreate the excitement of a seasoned scholar searching for just the right talent for the fashioning of a pure research topologist. R. L. Moore is such a scholar in action. "Nobody can possibly teach by the Moore-Method other than R. L. Moore himself," a former student once said. "The thing that makes him so successful and causes him to get so much work out of his students is that he knows just what to say and when to say it and how to say it so that it stimulates the very best effort in each individual student. When a good student shows up in one of Moore's classes, Moore is immediately challenged to see whether he can make that student perform. He watches the student, watches his reactions, and, upon seeing the first sign of understanding, he knows just how to coax that out and turn it into a performance. Then, you see, he has the student spending time on Moore's course." All are free to participate, and Dr. Moore's encouragement, support, and manipulation of students, is of the highest order. The following description by R. H. Bing may convey the depth and flavor of the involvement: "Moore used a great many techniques for teaching. I think he had a certain charm about his personality that attracted students to him. Probably none of Moore's students inspire the loyalty among students that Dr. Moore has engendered among his students. I don't believe that we have the same charm. Our students might like us. We may do a good job of training in mathematics but we don't have the same discipleship. He not only taught in the classroom but he taught in the halls. He taught through the influence of his wife. . . . Also I remember walking to school on a number of occasions with other professors and they would say, 'Moore has been telling me about the theorem that you have proved in class and he says that he is very impressed with you.' This is one of the techniques he had of encouraging his students by talking about them to other people and letting the word come back to them."

Immediately above the calculus, *Introduction to the Foundations of Analysis* carries the prerequisite: *Six hours of Calculus and consent of the instructor.* A student seeking a large body of information or expecting to be a passive member of a student group had best avoid the course. Lectures are not presented. Fundamental assumptions and definitions are introduced as needed. Theorems to be

proved are stated. No text is used. Professor Moore has never written a text based on this course nor is he avid to have others do so. Leading the student through unfamiliar material, presented without the aid of printed matter is the procedure. The content of the course is not fixed; the procedure is. Coverage of material and its storage as a memory bank is not the objective. At first, only one-step reasoning is required to progress from assumptions or previously established results to the conclusion in the desired theorem. Usually several theorems are stated at once, numbered in the order of increasing difficulty, showing interdependence, and perhaps necessitating the discovery of and the definition of a property. Progress is made by the class. Someone may see that he can prove Theorem 3 provided Theorem 1 is true. His proof and its dependency on Theorem 1 will be examined. Inevitably false reasoning is presented upon occasion; this is turned to constructive use and the class is expected to develop a critical sense that will expose any mistakes in proofs presented. Thus give and take goes on among the students.

As the course advances through the year more difficult theorems are stated. Multiple step reasoning becomes necessary; the usefulness of lemmas becomes apparent. More time, more thought, more searching of previously presented material is necessary. Certain theorems stand out in Professor Moore's mind and he may say "I'm pleased you got Theorem 15." He might point out the discovery of a concept, saying, "You have singled out the idea of upper-semi-continuity." Also, from time to time false conjectures are introduced. Thus the student's critical ability is further developed. Search for counterexamples, when proofs break down repeatedly, becomes routine procedure. In the more advanced courses he will toss in unsolved problems, not making clear to the student that they are real research questions. Frequently when he raises such problems he does not know if the conjecture is correct or incorrect. One former student relates: "The first piece of research that I did under R. L. Moore, he tossed out one of these things and I got it. He made me go over the proof a second time in class. Finally, he said 'You come in this afternoon and let's go over that proof again.' We spent about an hour checking it over carefully and at the end of that time he said 'You write that up.' "

Moore's use of the axiomatic method differs significantly in intent and in consequence from the often encountered "development of the real number system." The introduction of a system of axioms that underlie familiar notions and operations is fascinating to mathematicians of some maturity and may deepen their insight. Such a set of axioms can often be understood by less able and even public school students. The wide use of such material is amply justified and if a student is subjected repeatedly to an axiomatic approach to various parts of mathematics through textbook presentation he comes eventually to accept, appreciate, and understand the fundamental role of axioms. The beauties of mathematical structure become more apparent to him, and advantages of an orderly approach to even nonmathematical material become better understood, leading some people to speak of Systems Analysis in a broad sense. But the

Moore-Method develops rugged individualism, as one man said. Indeed there are business men who attribute their success to the training they received in one or two courses with Professor Moore. Others think the desire to participate, the need to be unique, self-confident, self-directed, and to communicate, as well as the need for spontaneity are met. One Ph.D. recipient said, "I felt a great part of it was that you started from scratch and you built up an entity, in which you knew every piece." Under the Moore-Method the student is presented with a set of axioms based on a certain number of undefined terms from which *he* must create an understanding of the space or system set forth by the axioms. No models or edifying examples are presented. Imagination and conjecture must seek and give meaning to the logical inference of the axioms. This is in accord with Moore's statement: "That student is taught the best who is told the least."

Freedom for independent action is accorded each student, so long as he does not present another's work as his own in class. Long before unlimited cuts became a student fetish, members of Professor Moore's advanced classes were advised that if an individual did not wish to hear the proof of a particular theorem he was at liberty to leave the room. The presumption was that no student would avoid hearing the presentation of a proof except for the express purpose of preserving for himself the opportunity of creating his own proof. A spontaneous, informal honor system springs up among the students. To paraphrase a former student: "The work is a very personal thing. It is put on a level that you do not find subjects taught. Subjects are usually taught like mountains afar off, but when you study with Dr. Moore it is as a participant in a task at hand. When you are in his class you are talking about mathematics with him and with the students around. He has a personal enthusiasm about the subject that is very catching, everyone seems to have it. He obviously enjoys mathematics, does mathematics, and teaches mathematics. The students see that immediately. It is an enthusiasm that, even though a student may have been away from the University of Texas for some years, can still be recalled vividly. He gives one the feeling that mathematics is more than just a way of making a living; it is a way of life, an orderly fashion in which you will want to consider all things. It is a method of self-examination, of self-discovery. In his classes you are personally responsible for finding a proof; it is a question of whether or not you have the ability; consequently, discussing the problem with someone else would not be the way it should be done. It is a challenge he can make; this is something about his personality."

Imposing your superior knowledge upon others becomes tabu. A former student relates the following story. He and another long term student wished to join a summer course in the *Foundations of Geometry* which was being conducted for high school teachers, using a rigorous set of axioms, to enrich their understanding. They were admitted on the condition that they sit on the back row and not say anything until they were called on. "This is a classical example of the Moore-Method. He gave them a few axioms and some theorems to prove.

They got to one theorem and one of the high school teachers got up to give a proof and it had a flaw in it. I sat there and pressed my lips together and Moore looked as if he believed it completely. Then he said, 'Do we all understand that?' Everyone said yes. Then he said, 'Well, just to play safe, let's go through it again.' This man went through it a second time. Again Moore asked if they understood it. 'Yes,' they said. About that time I couldn't stand it any longer and I called out, 'Dr. Moore,' and he jumped right down my throat. Well it turned out that it could not be proved on the basis of the axioms already stated. He did not want me to spoil his game of making them discover this for themselves. You see the beautiful way is to get the class to understand the need for a new axiom before you give it to them. Occasionally he would call on us for a theorem that they couldn't prove but otherwise we were to keep our mouths shut." The Moore-Method, then, seems to attain the objectives so often set forth of (1) getting the student involved, (2) leading the student into new personal discoveries, and (3) letting each student proceed at his own pace.

"I think that Moore is a very fine teacher but I don't want to give credit for his being a very fine teacher to this method. I think the credit belongs to the man," says Eldon Dyer. Again and again his former students emphasize the fact that not every class is filled with the proof of theorems. Logic, its use, and its relation to mathematics is explored. Great emphasis is placed on the precise use of language. Clippings from newspapers and magazines are sometimes brought into the class. The class is asked to analyze what these clippings say and what they are meant to say. It may be this precise way of looking at nonmathematical ideas wields great influence on a student's philosophy, on his way of attacking things that concern him. It has been pointed out that Dr. Moore came in contact with eighty to one hundred students a year over more than a forty year period. Thus he came in contact with many nonprofessional mathematicians and there is considerable evidence he influenced a large number of them significantly. Others have pointed out that he is a southern gentleman. Students are invited into his home where he shares with them his love of classical music and subjects them to mild pranks such as measuring their height with an ordinary looking yardstick that is an inch short of three feet. At the university he generously spends time informally with members of his classes but he never permits undue familiarity or disrespect. He finds indirect and often humorous ways of showing his disapproval of certain kinds of conduct. Repeatedly, former students refer to his patience and assert that they are unable to equal him in this respect. Moreover, it has been noted that there is a tendency now to bring women into mathematics in greater numbers but R. L. Moore accepted Ph.D. candidates when talent seemed evident. Anna Mullikin, who received her Ph.D. in 1922, was the first woman to write her thesis with him. Among others are Lida K. Barrett, Mary Elizabeth Hamstrom, and Mary Ellen Rudin. His unusual physical vigor tends to appeal to the men in his classes. Eyewitnesses say that when he was about sixty-seven or sixty-eight years old there was a student who seemed to be taking up too much time bragging about

his physical feats. Professor Moore offered to show this young man what he could do; he did some impressive one arm push-ups. He did them with his feet up in a chair so that his head was angled down toward the floor; he did seven of them and jumped back up onto his feet without losing his wind. The student got the message. Professor Moore likes to box. The story persists that the first person he looked up when he went to Princeton as the first native American mathematician to tour on the *American Mathematical Society Visiting Lectureship* was the boxing coach, who was his sparring partner during the early days of his teaching career as an instructor at Princeton. Indeed it has been said that one of the secrets of his success as a teacher is that he is not afraid of a fight, either a physical one or a scholarly one. More to the point perhaps, he seems to inspire others with a feeling of their own mental strength, something quite different from being led to think that the professor is wonderful.

Perhaps the fact that Moore threw his students into competition with each other in his graduate courses resulted from his being a scrapper. "It was an enjoyable kind of competition; there was no bitterness about it; there was this feeling that you were being developed, that you were being prepared to be a research mathematician. There was a feeling that you were already a participant in the making of mathematics." W. L. Ayres puts it thus: "I have seen student after student go through for their degrees and usually the very first original work they do is their thesis. How much they do of that is somewhat doubtful at times. But under Moore you have had a long hard-knocks trail of research before you get to the doctor's thesis. His students are thus prepared to stand on their own feet, which is an advantage. They have done research for three years." Mary Ellen Rudin says: "I met Dr. Moore at age sixteen . . . He signed me up to take trigonometry from him because that was the course I already knew that I wanted to take . . . I started absolutely a blank sheet, and I ended up a producing mathematician. I took a course every year with Dr. Moore from the time that I entered as a freshman until I graduated with a Ph.D. I had tremendously good competition as a graduate student . . . The student isn't reading a book every day. He isn't doing the sort of homework that one does in the usual course. His work is to produce mathematics every day and by producing mathematics every day he will form a habit. When he gets a Ph.D. he will continue to 'produce mathematics everyday.' " Yet Burton Jones has said: "My reaction was that it was the problems that Dr. Moore had given me to work. There is such challenge in trying to solve these problems or putting together the parts you know to see if you can not learn a little something more about the situation. I think this is the real power of the method. If you are mechanically inclined, you might compare it with having come upon a clock, say, that someone else has taken apart and you wonder if you can put it back together. There are some people that can't resist this."

When Dr. Moore's students go to other universities as faculty members and seek to talk about mathematics with their colleagues they do not find the same paternal encouragement existing among the older members of the faculty.

This is a shock. Suddenly they see the need for a broad spectrum of known mathematical results so that communication among fellow faculty members will be fun again. Most of them spend the next few years in self-education, broadening their view of the subject. Prime examples of this are R. L. Wilder, G. T. Whyburn, and R. H. Bing. One might say they climb those "mountains afar off" and having done so introduce new vistas. They have introduced new cross connections between differing branches of mathematics, and sought new ways of attacking old problems. They have stimulated their colleagues through their eagerness to learn, to create, and to teach. They have contributed to American mathematics not only worthwhile research treatises, Colloquium Lectures of the American Mathematical Society, expository papers, books coordinating and making available current mathematical research, but they have also served as Presidents of the American Mathematical Society and the Mathematical Association of America, as editors of journals, as members of boards and committees for the federal government, and as participants in the moves to improve mathematics teaching at all levels. They have tried to search out and use the most fruitful portions of the Moore-Method and even to introduce this method to all mathematicians since the method seems to attract students into mathematics and increase the ability of nonmathematicians to apply mathematical methods to their problems.

It has often been said that Moore dominates the Department of Mathematics at the University of Texas, and that graduate students spend a disproportionate part of their time on Moore's courses. One former student has said, "The year I took *Foundations of Analysis Situs*, I took a course in *Potential Theory* in which I fear I showed very little potential." However, the second and third generations of Moore-Method practitioners have not found it necessary for their students to neglect their other courses and have functioned well in large departments where the Ph.D. in other fields is awarded in considerable numbers. Each of these professors modifies the method, adapting it to his immediate surroundings. At one institution where the beginning graduate course is based on a set of axioms either for a semimetric space, a Moore space, or an abstract space, the students are asked not to read for the Fall term. It is made clear that the objective is to develop the ability to find proofs for one's self. By the end of the Fall term, however, they are encouraged to buy and begin to read J. L. Kelly's book on *General Topology* and they are warned that half of the examination in the Spring will be on this outside reading. Meanwhile, the class resumes personal development of proofs through the study of continua, transformations, or some other area not specifically treated in Kelly. At one university, the beginning graduate course in topology begins with a set of axioms, proceeds in the Moore manner accommodating the pace to the particular student group until the end of the first semester, then uses a mixed array of theorem proving sessions interspersed with lectures. At another university any one of several professors may conduct the first course in topology, each using some variation of the Moore-Method. During the student's second year there are

several courses available in different areas of topology from which he may choose one or two classes. The average time for a Ph.D. in this department being five years, he is able to acquaint himself with a relatively wide spectrum of topology as well as develop his creative ability.

"The Moore-Method is a geometric method with an adequate set of axioms and the theorems which he asks his students to prove are just like the 'originals' proposed in some geometry texts. The difference is merely a matter of sophistication or style." "The Moore-Method is a modification of the German seminar where the basic objective was affording the professor and student the opportunity to do mathematics together." "The Moore-Method is the Socratic method." "The Moore-Method is the man himself." Many mathematicians understand the Moore-Method but putting it into a concisely worded definition seems to vary. Perhaps we had best take it as our undefined notion and simply try to describe some of its characteristics. R. L. Moore was the first mathematician to write his thesis for his Ph.D. with Oswald Veblen, according to the listing in the *American Mathematical Society Semicentennial Publications, vol. 1* (*History*), A Ph.D. recipient from Princeton University has said of Professor Veblen, "Without having too good a grasp of a subject, he could chat about it and stimulate others to produce in it." Getting students to produce mathematics is certainly one of Professor Moore's main objectives. To quote from the *AMS Semicentennial History*, "R. L. Moore's first publication was a neat redundancy proof which he discovered while a graduate student in 1902, and which Halsted published." George Bruce Halsted was Professor of Mathematics at the University of Texas at the time. Giving a student some chance to consider unresolved mathematical questions and produce new mathematics early in his graduate experience is habitual with Professor Moore. The above paper put him in correspondence with E. H. Moore and he later studied at the University of Chicago, where E. H. Moore was the first Chairman of the Mathematics Department after the founding of the University of Chicago, and where he had invited Oskar Bolza and Heinrich Maschke to become members of his department. "Moore was a fiery enthusiast, brilliant, and keenly interested in the popular mathematical research movements of his day; Bolza, a product of the meticulous German school of analysis led by Weierstrass, was an able and widely read research scholar; Maschke was more deliberate than the other two, sagacious, brilliant in research, and a most delightful lecturer . . . These three men supplemented one another remarkably . . . During the period of 1892–1908 the University of Chicago was unsurpassed in America as an institution for the study of mathematics." Other remarks about E. H. Moore from the *AMS Semicentennial History* reveal: "In the lecture room Professor Moore's methods defied most established rules of pedagogy. Such rules, indeed, meant next to nothing to him in the conduct of his advanced courses. He was absorbed in the mathematics under discussion to the exclusion of everything else . . . " Definitely, this description points to perceptual influence on R. L. Moore by E. H. Moore. Further evidence of this might lie in the fact that Professor E. W. Chit-

tenden was invited to the University of Texas to teach a graduate course in the summer and that he used a similar method.

Under the sponsorship of the Mathematical Association of America some effort has been made to demonstrate the Moore-Method in Summer Institutes for Teachers and to provide visiting lecturers who sometimes talk about the Moore-Method and attempt to widen the understanding of topology. Professor Burton Jones of the University of California at Riverside has experimented with the method in Australia and lectured on it in various Indian universities. Professor R. H. Bing has generalized the method and is using it to arouse a group attack on the Poincaré conjecture. Professor E. E. Moise at Harvard University has established a program for teachers which reflects the fact that he worked with Professor Moore for his Ph.D. Also, considerable work is being done in the School of Education at the University of Wisconsin that is related to the Moore-Method. In the opinion of the author, at least, R. L. Moore will not begrudge anyone the use of his method even in modified form so long as its use effectively develops in a student creative mathematical ability, the correct use of logical inference, or careful precise expression of ideas.

The author wishes to express her gratitude for the ideas, suggestions, and encouragement accorded her through letters and personal interviews by the following mathematicians: W. L. Ayres, B. J. Ball, Gayle Ball, R. H. Bing, W. L. Duren, Eldon Dyer, H. M. Gehman, George Henderson, F. Burton Jones, J. H. Roberts, Mary Ellen Rudin, J. M. Slye, J. M. Thomas, F. P. Welch, W. M. Whyburn, J. M. Worrell, and Leo Zippin. She is also indebted to her husband, Gordon T. Whyburn, for his kindly criticism and helpfulness.

AN ISOPERIMETRIC PROPERTY OF SURFACES WITH MOVEABLE BOUNDARIES

J. C. C. NITSCHE, Institute of Technology, University of Minnesota

Viewing various experiments with soap films, one is led to consider the following generalization of Plateau's problem:

Determine a surface of a prescribed topological type and of least area which is bounded by a configuration of several fixed contours, supporting surfaces, and moveable threads of given lengths.

For a list of pertinent literature see [**15**], page 239. Soap laminae spanning frames with moveable contours—experimentally realized by hairs or thin silk threads—have been described, among other places, in [**3**], p. 48; [**4**], p. 174; [**14**], p. 584.

While Plateau's problem is always solvable, it is a well-known fact (for literature see [**18**]) that the existence of a solution for the general problem hinges on certain metric-geometric conditions which must be satisfied by the bounding configuration. Otherwise the minimizing sequences on which the exis-

tence proof is based do not converge or, if they do, converge only to degenerate limit structures—disconnected surfaces of lower topological types or one-dimensional formations. Two simple examples may illustrate these contingencies: A pair of distinct Jordan curves at too large a distance is not capable of bounding a doubly-connected minimal surface. A frame consisting of a fixed Jordan arc in space and a moveable thread of equal length with the same end points cannot span a nondegenerate surface of least area, since this surface would collapse into the Jordan arc.

In a number of cases, however, existence proofs are feasible by extensions of the classical methods. The boundary behavior of the solution surface is fully studied for the fixed contours ([8], [10], [12], [17]) and satisfactorily understood for the free boundaries, i.e., the boundary portions on supporting manifolds ([5], [6], pp. 218–223, [9], [13]). Nothing is known yet for the moveable contours. For well-posed problems it is very likely that the solution surface will be analytic up to the moveable parts of the boundary and free of branch points on these parts. Anticipating this fact, i.e., assuming the existence of a solution surface with sufficient regularity properties, we shall derive here a geometrical property which characterizes those contours that are freely moveable, subject only to the fixation of their lengths and, possibly, their end points.

THEOREM. *The moveable parts of the boundary must be curves of constant space curvature. On the solution surface itself they are asymptotic lines of constant geodesic curvature.*

Variational problems describing surfaces with variable boundaries have been touched upon in the literature before (see, in particular, [2], p. 671). It is curious to note, however, that so far no results of a geometric nature seem to have been extracted from the formal conditions. There is an interesting connection to the isoperimetric problem on surfaces which has been extensively treated over the years since it was first discussed by F. Minding and J. Steiner. For the newer literature on this subject consult the bibliography in [16], for the older literature compare [11], pp. 614–620; see further [1], p. 154 and [7], pp. 149–153. If a moveable contour were to vary on a fixed surface, its optimal position would have to be that of a curve of constant geodesic curvature. For the problem at hand, of course, the solution surface itself is unknown at the start. Considering that closed curves of constant geodesic curvature very possibly may not exist in certain regions of a (fixed) surface, one realizes that our general problem still poses many questions and invites interesting comparisons, particularly in cases where freely moveable contours are provided.

We shall now prove our theorem for the concrete case of a frame which consists of a fixed rectifiable Jordan arc Γ_1 and a moveable arc Γ_2 of prescribed length (smaller than that of Γ_1) having its end points in common with Γ_1. Denote by P the semidisc $\{u, v; u^2+v^2<1, v>0\}$ and let $S=\{\mathfrak{x}=\mathfrak{x}(u, v); (u, v)\in\overline{P}\}$ be a surface of least area bounded by this frame. It is assumed here that the position vector of S possesses the following properties:

(i) $\mathfrak{x}(u, v)$ belongs to class $C^0(\overline{P})\cap C^2(P)$ and satisfies in P the relations $\Delta\mathfrak{x}=0$, $\mathfrak{x}_u^2=\mathfrak{x}_v^2$, $\mathfrak{x}_u\mathfrak{x}_v=0$.

(ii) $\mathfrak{x}(u, v)$ provides a topological mapping of the boundary portions $\partial_1 P = \{u, v;\ u^2+v^2=1,\ v\geqq 0\}$ and $\partial_2 P=\{u, v;\ -1\leqq u\leqq 1,\ v=0\}$ onto the curves Γ_1 and Γ_2, respectively.

(iii) $\mathfrak{x}(u, v)$ is analytic (or at least sufficiently regular) in P as well as in the interior points of $\partial_2 P$, and $\mathfrak{x}_u^2>0$ in the latter points. For $0<r<1$ we set $P_r=\{u, v;\ u^2+v^2<r^2,\ v>0\}$ and we define the boundary portions $\partial_1 P_r$ and $\partial_2 P_r$ as before. S_r then will denote the portion $\{\mathfrak{x}=\mathfrak{x}(u, v);\ (u, v)\in\overline{P}_r\}$ of the surface S.

Now let $\mathfrak{z}(u, v)$ be an arbitrary vector which is sufficiently regular in P_r, $0<r<1$, vanishing along $\partial_1 P_r$, and consider the varied surface

$$S_r(\epsilon) = \{\mathfrak{x} = \mathfrak{y}(u, v;\ \epsilon);\ (u, v) \in \overline{P}_r\}$$

whose position vector is $\mathfrak{y}(u, v;\ \epsilon)=\mathfrak{x}(u, v)+\epsilon\mathfrak{z}(u, v)$. In order for $S_r(\epsilon)$ to be admissible, the length of the image of $\partial_2 P_r$ under the mapping by $\mathfrak{y}$ cannot depend on ϵ. Remembering the minimizing property of S we are thus led to a variational problem with side conditions: The area $A[S_r(\epsilon)]$ must have a stationary value for $\epsilon=0$ subject to the condition $\int_{-r}^{r}|\mathfrak{y}_u(u, 0;\ \epsilon)|\,du=\text{const.}$

Using the relation $\mathfrak{x}_u\mathfrak{z}_u+\mathfrak{x}_v\mathfrak{z}_v=(\partial/\partial u)(\mathfrak{x}_u\mathfrak{z})+(\partial/\partial v)(\mathfrak{x}_v\mathfrak{z})$ and applying Green's theorem we find

$$\begin{aligned}
A[S_r(\epsilon)] &= \frac{1}{2}\iint_{P_r}(\mathfrak{y}_u^2 + \mathfrak{y}_v^2)\,du\,dv \\
&= A[S_r(0)] + \epsilon\iint_{P_r}(\mathfrak{x}_u\mathfrak{z}_u + \mathfrak{x}_v\mathfrak{z}_v)\,du\,dv + \epsilon^2[\cdots] \\
&= A[S_r(0)] - \epsilon\int_{-r}^{r}\mathfrak{x}_v(u, 0)\mathfrak{z}(u, 0)\,du + \epsilon^2[\cdots].
\end{aligned}$$

Furthermore

$$|\mathfrak{y}_u(u, 0;\epsilon)| = |\mathfrak{x}_u| + \epsilon\frac{\mathfrak{x}_u\mathfrak{z}_u}{|\mathfrak{x}_u|} + \epsilon^2[\cdots] + \cdots.$$

By the rules of the calculus of variations (see, for instance, [2], pp. 457–465) we can conclude that

$$\int_{-r}^{r}\left[\mathfrak{x}_v\mathfrak{z} + \lambda\frac{\mathfrak{x}_u\mathfrak{z}_u}{|\mathfrak{x}_u|}\right]_{(u,0)}du = \int_{-r}^{r}\left[\mathfrak{x}_v - \lambda\frac{\partial}{\partial u}\left(\frac{\mathfrak{x}_u}{|\mathfrak{x}_u|}\right)\right]_{(u,0)}\mathfrak{z}(u, 0)\,du = 0.$$

Here λ is a (constant) multiplier. Since $\mathfrak{z}(u, v)$ was an arbitrary vector the relation

$$\mathfrak{x}_v(u, 0) = \lambda\frac{\partial}{\partial u}\left(\frac{\mathfrak{x}_u(u, 0)}{|\mathfrak{x}_u(u, 0)|}\right)$$

follows. Introducing the arc length on Γ_2, $s=s(u)=\int_{-1}^{u}|\mathfrak{x}_u(u,0)|\,du$, and denoting the differentiation with respect to s by a prime, this relation becomes

$$\lambda\mathfrak{x}''(u,0)=\frac{\mathfrak{x}_v(u,0)}{|\mathfrak{x}_u(u,0)|}\quad\text{for } -r\leqq u\leqq r.$$

Since the vector on the right hand side is of unit length, while $|\mathfrak{x}''|$ denotes the curvature of Γ_2, the first part of our theorem follows. The second part of the theorem is a consequence of the fact that $\mathfrak{x}_v$ is a tangent vector on S. Thus, if Γ_2 is considered as a curve on S, its normal curvature must vanish. Such a curve is an asymptotic line.

Acknowledgment. The preceding work was sponsored by the Air Force Office of Scientific Research under AFOSR Grant No. 883-67.

References

1. W. Blaschke, Vorlesungen über Differentialgeometrie I, 4th ed., Springer, Berlin, 1945.
2. O. Bolza, Vorlesungen über Variationsrechnung, Teubner, Leipzig and Berlin, 1909.
3. C. V. Boys, Soap bubbles: Their colours and the forces which mold them, Dover, New York, 1959.
4. R. Courant, Soap film experiments with minimal surfaces, this MONTHLY, 47 (1940) 167–174.
5. ———, On Plateau's problem with free boundaries, Proc. Nat. Acad. Sci. U.S.A., **31** (1945) 242–246.
6. ———, Dirichlet's Principle, Conformal Mapping, and Minimal Surfaces, Interscience, New York, 1950.
7. G. Darboux, Leçons sur la théorie générale des surfaces et les applications géométriques du calcul infinitésimal, 3rd vol., Gauthier-Villars, Paris, 1884.
8. E. Heinz and F. Tomi, Zu einem Satze von S. Hildebrandt über das Randverhalten von Minimalflächen, Math. Z., 111 (1969) 372–386.
9. S. Hildebrandt, Über Minimalflächen mit freiem Rand, Math. Z., 95 (1967) 1–9.
10. ———, Boundary behavior of minimal surfaces, Arch. Rational Mech. Anal., 35 (1969) 47–82.
11. A. Kneser, Variationsrechnung, Enc. d. math. Wissensch., 2.1.1, Teubner, Leipzig, 1899–1916, pp. 571–625.
12. H. Lewy, On the boundary behavior of minimal surfaces, Proc. Nat. Acad. Sci. U.S.A., 37 (1951) 103–110.
13. ———, On minimal surfaces with partially free boundary, Comm. Pure Appl. Math., 4 (1951) 1–13.
14. H. Minkowski, Kapillarität, Enc. d. math. Wissensch., 5.1.9, Teubner, Leipzig, 1903–1921, pp. 558–613.
15. J. C. C. Nitsche, On new results in the theory of minimal surfaces, Bull. Amer. Math. Soc., 71 (1965) 195–270.
16. ———, The isoperimetric inequality for multiply-connected minimal surfaces, Math. Ann., 160 (1965) 370–375.
17. ———, The boundary behavior of minimal surfaces—Kellogg's theorem and branch points on the boundary, Inventions Math., 8 (1969) 313–333.
18. ———, and J. Leavitt, Numerical estimates for minimal surfaces, Math. Ann., 180 (1969) 170–174.

AN APPROACH TO EMPIRICAL LOGIC

C. H. RANDALL AND D. J. FOULIS, University of Massachusetts

1. Introduction. The aim of an empirical science is to order, explain, and predict the observable consequences of certain physical operations. The mathematical apparatus for such an empirical science, therefore, ought to be based on a formal "empirical logic" capable of describing these events. Such logics, in turn, ought to be erected on a refinement of the conventional notion of a physical operation. A rigorous development, based on this desideratum, leads to a nonclassical symbolic logic that bears a striking resemblance to the so-called logic of quantum mechanics [2, 3, 6, 8, 15, 18, 19, 20, 25, 27, 32, 34, 35, 36, 37, 39, 44, 45, 48, 49, 50].

By an *operation* we shall mean a collection of instructions that describe a well-defined physically realizable (reproducible) procedure, and that in addition specifies what must be observed and recorded. In particular, an operation must require that, as a result of every execution of the instructions, one and only one symbol from a specified set R be recorded as the outcome of that realization of the operation. As a consequence of this definition, R is an exhaustive and mutually exclusive set of outcomes for the corresponding operation. Carefully note, if we delete or add details to such a set of instructions, and in particular if we modify the outcome set R in any way, we thereby define a *new* operation.

Here, we are by no means restricting our attention to operations involving only traditional laboratory procedures. In fact, test procedures on an assembly line, data gathering procedures in general (such as opinion polling), pencil-and-paper procedures (such as executing computational algorithms), and even procedures involving subjective approvals or disapprovals are all admissible provided that they satisfy the above criteria.

This definition appears to be the only tractable one, since the only means of establishing that two individuals performed the same operation is with a description. Niels Bohr repeatedly made this point in connection with quantum physics. He said [4, p. 3]:

> "The decisive point is to recognize that the description of the experimental arrangement and the recording of observations must be given in plain language, suitably refined by the usual physical terminology. This is a simple logical demand, since by the word 'experiment' we can only mean a procedure regarding which we are able to communicate to others what we have done and what we have learnt."

2. Compositions of operations. In practice, operations are normally synthesized from sequences of more primitive procedures and arrangement ap-

C. H. Randall received his Ph. D. at Rensselaer Polytech in 1966 under C. Lemke. He has held positions at G. E. Knolls Atomic Lab., Martin Co., Westinghouse, and Univ. of Massachusetts. His main research interest is in the foundations of empirical science.

David Foulis received his Ph. D. at Tulane in 1958 under F. Wright. He has held positions at Lehigh, Wayne State, Univ. Florida, and Univ. of Massachusetts. His main research is in abstract algebra and empirical logic, and he has published *Fundamental Concepts of Mathematics* with Prindle, Weber, & Schmidt, 1969. *Editor.*

provals, however informal or subtle. Our first task will be to obtain a formal representation for such a synthesis. By a *primitive operation*, we mean roughly an operation which we cannot, or do not choose to, factor into sequences of more basic operations. (In the final analysis, the decision to regard an operation as being primitive must be largely subjective.)

Let $\mathcal{P}$ be a set of primitive operations. It will be mathematically convenient to assume that, for each operation $\omega \in \mathcal{P}$, the outcome set R_ω for ω contains at least two distinct elements. This is no real restriction, since we can always modify an operation in an essentially inconsequential way by appending a fictitious additional outcome to its outcome set.

There is no guarantee that the various outcome sets R_ω are mutually disjoint. In order to obviate this difficulty, we shall simply replace each R_ω by the *labeled outcome set* $X_\omega = \{\omega\} \times R_\omega$. Clearly, this is an inconsequential modification, for we are only agreeing to record an outcome of the primitive operation ω as (ω, b) rather than simply as b. We define $X = \bigcup \{X_\omega | \omega \in \mathcal{P}\}$, and we call X the *labeled outcome space* for $\mathcal{P}$. If $x = (\alpha, a), y = (\beta, b) \in X$, we say that x and y are *orthogonal* and we write $x \# y$ if and only if $\alpha = \beta$ and $a \neq b$. Thus, two labeled outcomes in X are orthogonal precisely when they are distinct outcomes for the same primitive operation in $\mathcal{P}$.

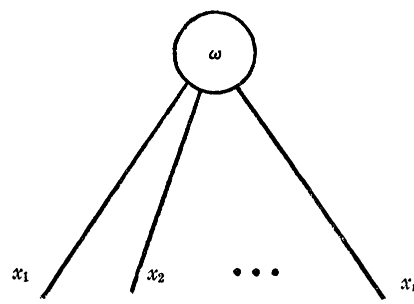


FIG. 1

A given primitive operation $\omega \in \mathcal{P}$, together with its associated labeled outcome set X_ω, can be conveniently depicted as shown in Figure 1, in which the labeled circle represents the operation ω and the labeled "legs" represent the outcomes $x_1, x_2, \cdots, x_n$ comprising X_ω. The diagram in Figure 1 will be called the *operational diagram* for ω.

Compound operations, synthesized from sequences of the primitive operations in $\mathcal{P}$, can now be depicted by a natural generalization of Figure 1. For instance, see Figure 2, which depicts a compound operation E for which the primitive operation α is the initial step. In the event that the labeled outcome x_1 is obtained when this first step α is executed, the primitive operation β is the required second step. In the event that the labeled outcome x_2 is obtained when α is executed, the primitive operation γ is the required second step, etc., etc.; and likewise for the third steps. Notice, however, that if γ is the required second

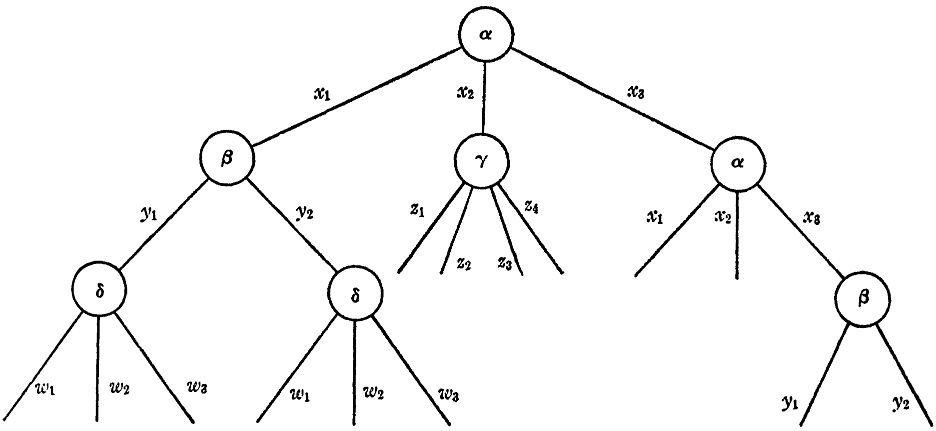


FIG. 2

step, no third step is specified and the operation E is terminated. The diagram in Figure 2 will be called the *operational diagram* for the compound operation E.

Suppose that the compound operation E of Figure 2 is executed and that x_1 is the labeled outcome obtained when the first step α is executed, y_1 is the labeled outcome obtained when the second step β is executed, and w_1 is the labeled outcome obtained when the third step δ is executed. We shall then agree to record the *outcome* of this execution of E as the formal sequence $x_1y_1w_1$. For the compound operation E, there are altogether fourteen possible outcomes, so that the outcome set R_E is given by

$$R_E = \{x_1y_1w_1,\ x_1y_1w_2,\ x_1y_1w_3,\ x_1y_2w_1,\ x_1y_2w_2,\ x_1y_2w_3,\ x_2z_1,\ x_2z_2,\ x_2z_3,\ x_2z_4,\\ x_3x_1,\ x_3x_2,\ x_3x_3y_1,\ x_3x_3y_2\}.$$

Evidently, the operational diagram for the compound operation E can be reconstructed merely from a knowledge of the outcome set R_E. For example, all of the formal sequences in the set R_E initiate with x_1, with x_2 or with x_3, making it clear that the initial operation for E is the unique primitive operation α for which $X_\alpha = \{x_1, x_2, x_3\}$. An obvious inductive procedure yields the complete operational diagram for E.

Let S denote the set of all finite formal sequences of elements of the labeled outcome space X. A typical element $a \in S$ will have the form $a = u_1u_2 \cdots u_n$, with $u_1, u_2, \cdots, u_n \in X$. We define the *length* of a, in this case, by $\text{length}(a) = n$. If we define a "multiplication" operation on S by

$$(u_1u_2 \cdots u_n)(v_1v_2 \cdots v_m) = u_1u_2 \cdots u_nv_1v_2 \cdots v_m,$$

then S becomes a semigroup; in fact, S is just the free semigroup over the set X. We identify the elements of length 1 in S with the corresponding elements of X, so that X is a subset of S. It will be mathematically convenient to consider the

monoid (semigroup with unit) obtained from the semigroup S by formally adjoining a unit element 1 to S and defining $a\cdot 1=1\cdot a=a$ and $1\cdot 1=1$ for all $a\in S$. We denote the resulting monoid by $\Gamma=\Gamma(\mathcal{P})=S\cup\{1\}$, and we define length $(1)=0$.

A subset M of Γ will be said to be *bounded* if there exists a nonnegative integer n such that length $(a)\leqq n$ for every $a\in M$. If M is a nonempty bounded subset of Γ, we define length $(M)=\max\{\text{length}(a)\mid a\in M\}$. The outcome set R_E corresponding to a compound operation E depicted by an operational diagram (such as Figure 2) is evidently always a bounded subset of Γ and length (R_E) gives the maximum number of sequential executions of primitive operations that might be required during an execution of the compound operation E.

Not every bounded subset of Γ is the outcome set for a compound operation E. In order to characterize those subsets of Γ that have the form R_E, we must introduce a new definition. We shall say that two elements a, $b\in\Gamma$ are (*lexicographically*) *orthogonal*, in symbols $a\perp b$, if and only if there exist elements $c, d, e\in\Gamma$ and elements $x, y\in X$ such that $x\# y$, $a=cxd$ and $b=cye$. If M is a subset of Γ, then we call M an *orthogonal set* if and only if a, $b\in M$ with $a\neq b$ implies that $a\perp b$.

It is not difficult to prove that a subset M of Γ is of the form R_E for some compound operation E if and only if $M\neq\{1\}$ and M is a bounded maximal orthogonal subset of Γ. In what follows, we shall regard the set $\{1\}$ as the outcome set of a trivial operation I requiring only that the symbol 1 be recorded.

An unbounded maximal orthogonal subset of Γ can be regarded as the outcome set of an "in principle" compound operation which might not be physically realizable because of the possibility that an infinite number of steps could be required for its execution. Despite the fact that "in principle" compound operations might fail to satisfy our requirement of physical realizability, they can be considered as "limiting cases" of physically realizable compound operations, and it is therefore reasonable to expect our mathematical formalism to be able to represent them.

As it has been interpreted above, an operational diagram can be viewed as a set of instructions. Since the operational diagram for a compound operation E is completely determined by its set of outcomes R_E, this set itself can be regarded as a complete description of the required procedure. Recalling that by definition an operation is precisely such a set of instructions, we propose to identify compound operations with their outcome sets. This is formally accomplished by the following definition: A *compound $\mathcal{P}$-operation* is a maximal orthogonal subset of $\Gamma(\mathcal{P})$. The set of all compound $\mathcal{P}$-operations is denoted by $\mathcal{E}=\mathcal{E}(\mathcal{P})$. The set of all bounded compound $\mathcal{P}$-operations is denoted by $\mathcal{E}_0=\mathcal{E}_0(\mathcal{P})$.

If $E\in\mathcal{E}$, it is natural to regard $D\subset R_E$ as a proposition asserting that E was executed and in this realization of E one of the outcomes $d\in D$ was recorded. This we recognize as precisely the Kolmogorov [23] point of view on which much of modern statistics has been founded. Following Kolmogorov, we shall also refer to such sets as *events*. It is not difficult to see that these events are

precisely the orthogonal subsets of Γ. The set of all events is denoted by $\mathcal{O}=\mathcal{O}(\mathcal{P})$ and the set of all bounded events by $\mathcal{O}_0=\mathcal{O}_0(\mathcal{P})$.

Notice that, in general, a given $D\in\mathcal{O}$ will be a subset of many operations $E\in\mathcal{E}$; hence, unlike Kolmogorov, we cannot associate such an event with only a single operation. Kolmogorov explicitly considered only one operation at a time, while we are simultaneously considering all of the operations in $\mathcal{E}$. For example, let ω, λ, γ, $\delta\in\mathcal{P}$ with $X_\omega=\{u, u'\}$, $X_\lambda=\{v, v'\}$, $X_\gamma=\{w, w'\}$ and $X_\delta=\{x, x'\}$. Then $E_1=\{xu, xu', x'v, x'v'\}$ and $E_2=\{xu, xu', x'w, x'w'\}$ are in $\mathcal{E}_0(\mathcal{P})$ and xu is an outcome of both operations. Furthermore, $\{xu, xu'\}$ is an event for both operations.

Our inability to assign a unique operation to each outcome in Γ, and consequently to each event in $\mathcal{O}$, is less of a predicament than it appears to be. In the above example, if we execute δ and record x and then execute ω and record u, we have in fact followed the instructions for both E_1 and E_2. It was carefully noted earlier that we would regard as a realization of an operation *any* execution of a physical procedure that satisfied the *explicit* specifications in the instructions; consequently, the recording of xu was simultaneously an outcome for both E_1 and E_2. The same argument applies to events. Thus, we need to modify the suggested Kolmogorov interpretation of an event by deleting any reference to a specific operation.

3. The calculus of events. The structure of our empirical logic will be crucially affected by our decisions concerning the assignment of truth values to the propositions represented by the events $D\in\mathcal{O}$. In the absence of a dogmatic enforcement of suitable stipulations concerning the verification of such propositions, we are left vulnerable to an insidious difficulty—the validity of evidence. Data is frequently challenged on the grounds that the experimental procedures that produced it did not properly account for ______. In practice, it is never difficult to fill in such a blank; the possibilities are literally inexhaustible. An objective line must be drawn somewhere, and in this section, we make our choice!

Notice that our failure to associate a unique operation $E\in\mathcal{E}$ with an outcome $b\in\Gamma$ amounts to rejecting, as admissible information, any knowledge of the operations that would have been executed if outcomes other than those recorded had resulted. In other words, any intention to execute one of these potential operations, and any unrecorded consequences of such an intention, are regarded as being inadmissible data. *Only the realizations of the outcomes of compound $\mathcal{P}$-operations are admissible as valid evidence in the empirical universe of discourse erected on $\mathcal{P}$.* The events $D\in\mathcal{O}$ are to represent the propositions that are validated or refuted by such evidence. We emphasize that, in our view, such a proposition is well-defined if and only if the exact conditions under which it is to be regarded as true, as well as those under which it is to be regarded as false, are stipulated in terms of valid evidence. Thus, we must assign to each $D\in\mathcal{O}$ a subset $D^+\subseteq\Gamma$ of outcomes for which D is regarded as validated and a subset $D^-\subseteq\Gamma$ of outcomes for which D is regarded as refuted.

A strict classical interpretation of an event D would require that we put $D^+ = D$ and $D^- = \Gamma \backslash D$. For reasons which we can only outline briefly here, we shall, however, define D^+ and D^- differently. We wish to place three requirements on the map $D \mapsto D^+$, namely:

(1) $ab \in \{a\}^+$ for $a, b \in \Gamma$.
(2) For $A, B \in \mathcal{O}$, $A \subseteq B \rightarrow A^+ \subseteq B^+$.
(3) For $a \in \Gamma$, $E \in \mathcal{E}$, $\{a\}^+ \subseteq (aE)^+$, where $aE = \{ae \mid e \in E\}$.

The motivation for requirements (1) and (2) is plain enough, but requirement (3) might be somewhat unexpected. In (3), we suggest that the event aE be regarded as asserting the validation of the event $\{a\}$ and a possible subsequent intention to carry out the operation E. Since we do not care to rule out the *potential* execution of any of the operations $E \in \mathcal{E}$, we are obliged to impose condition (3). Requirements (1), (2) and (3) generate the following, which is our official formal definition of D^+, for $D \in \mathcal{O}$: $D^+ = \{a \in \Gamma \mid \exists\, b, c \in \Gamma \text{ and } \exists\, E \in \mathcal{E} \text{ with } a = bc,\ bE \subseteq D\}$.

Our candidate for D^- is perhaps more transparent. We propose to view a realization of an outcome $e \in \Gamma$ as refuting any event claiming the occurrence of outcomes all of which are orthogonal to e. Hence, for $D \in \mathcal{O}$, we define $D^- = \{e \in \Gamma \mid \forall d \in D,\ e \perp d\}$.

Let $E \in \mathcal{E}$ and let $D \in \mathcal{O}$. If $D \subseteq E$, we shall say that the operation E is a *direct test* for the event D, since in this case each execution of E produces an outcome $d \in E$ for which we can directly test the validity of D (without reference to intended or potential operations). If $E \subseteq (D^+ \cup D^-)$, we shall say that the operation E *tests* D, since in this case each execution of E produces an outcome $d \in E$ for which we can still test the validity of D by checking whether or not d belongs to D^+.

The above considerations lead us to define the binary relation $\prec$ of *strict implication* on $\mathcal{O}$ as follows: For $C, D \in \mathcal{O}$, $C \prec D$ if and only if $C^+ \subseteq D^+$. As one might hope, it can be shown that $C \prec D$ if and only if $D^- \subseteq C^-$. The relation $\prec$ is reflexive and transitive on $\mathcal{O}$, but it is not in general antisymmetric. Hence, we are led to define the equivalence relation $\equiv$ of *strict equivalence* on $\mathcal{O}$ by $C \equiv D$ if and only if $C \prec D$ and $D \prec C$.

Notice that, for $b \in \Gamma$, $E \in \mathcal{E}$, we have $\{b\} \equiv bE$; but, in general, $\{b\} \not\equiv Eb$, where $Eb = \{eb \mid e \in E\}$. This "temporal asymmetry" is a consequence of our definition of lexicographic orthogonality and our decisions regarding valid evidence. An event bE has been viewed as asserting that $\{b\}$ has been validated, and essentially only that. On the other hand, Eb asserts more, since with regard to b the operation E is no longer potential, but realized. In a crude sense, the execution of E is viewed as exercising a possible influence on the outcomes of subsequent operations; whereas, in drawing our "objective line," we consciously failed to take into account the possibility of intended operations as exercising such an influence.

In addition to the relations of strict implication and strict equivalence, there are three other binary relations on $\mathcal{O}$ that are germane to our calculus of events. Let $C, D\in\mathcal{O}$. We define $C\leftrightarrow D$ if and only if $C\cup D\in\mathcal{O}$. We also define $C\perp D$ if and only if $C\leftrightarrow D$ and $C\cap D=\varnothing$. Finally, we define $C\perp_m D$ if and only if $C\perp D$ and $C\cup D\in\mathcal{E}$.

Notice that, for $C, D\in\mathcal{O}$, $C\leftrightarrow D$ if and only if there exists an operation $E\in\mathcal{E}$ such that $C\subseteq E$ and $D\subseteq E$; that is, $C\leftrightarrow D$ if and only if the two events C and D admit a simultaneous direct test operation $E\in\mathcal{E}$. There is, of course, a more general possibility; namely, that there exists an operation $E\in\mathcal{E}$ which simultaneously tests both C and D (but not necessarily directly). It is pleasant to be able to report that C and D are simultaneously testable in the latter sense if and only if there exist events $C_1, D_1\in\mathcal{O}$, such that $C_1\equiv C$, $D_1\equiv D$ and $C_1\leftrightarrow D_1$.

If $C\perp D$, we shall say that the event C is *orthogonal* to the event D. It is a theorem that $C\perp D$ if and only if $C^+\subseteq D^-$. Consequently, if $C\perp D$, then any validation of C constitutes at the same time an automatic refutation of D.

If $C\perp_m D$, we shall say that the event C is *maximally orthogonal* to the event D. It is a theorem that $C\perp_m D$ if and only if $C^+=D^-$. Consequently, if $C\perp_m D$, then an outcome validifies C if and only if it refutes D. Maximal orthogonality and strict equivalence are related by the following fact: The event C is strictly equivalent to the event D if and only if there exists an event B that is maximally orthogonal to both C and D.

In Kolmogorov's work [**23**], in which only one basic operation is under discussion, the events are represented by subsets of the sample space and their denials (or negations) are accordingly represented by the set theoretic complements of these subsets in the sample space. In our event calculus, a given event $D\in\mathcal{O}$ admits complements $E\backslash D$ in *any* of the operations $E\in\mathcal{E}$ for which $D\subseteq E$; hence, we can associate (in general) a multitude of relative negations $C=E\backslash D\in\mathcal{O}$ with our given D. Evidently, C is a relative negation for D in this sense if and only if $C\perp_m D$.

It is a theorem that the set of all events $C\in\mathcal{O}$ such that $C\perp_m D$ constitutes an equivalence class in $\mathcal{O}$ with respect to the relation of strict equivalence. This suggests that the equivalence class $\{C\in\mathcal{O}\mid C\perp_m D\}$ should be regarded, in some sense, as representing the absolute negation of the event D. This, in turn, suggests that we shift our attention from the *event calculus* $(\mathcal{O}; \prec, \equiv, \leftrightarrow, \perp, \perp_m)$ to the quotient structure $\mathcal{L}=\mathcal{O}/\equiv$ obtained from the event calculus by identification of strictly equivalent events.

4. The operational logic. If $D\in\mathcal{O}$, we define $[D]=\{C\in\mathcal{O}\mid C\equiv D\}$. Such an equivalence class will be referred to as an *experimental proposition*. We agree that the procedure for verifying $[D]$ is the same as that described above for verifying D. Evidently, this procedure is independent of the choice of the representing event D. We define $\mathcal{L}$ to be the set of all experimental propositions. For $[C]$, $[D]\in\mathcal{L}$, we shall say that $[C]$ *strictly implies* $[D]$, and write $[C]\leqq[D]$,

if and only if $C \prec D$. Clearly, $\leqq$ is a well-defined partial order relation on $\mathcal{L}$. Note that, for all $[D] \in \mathcal{L}$, we have $\{\varnothing\} = [\varnothing] \leqq [D] \leqq [\{1\}] = \mathcal{E}$; hence, $\{\varnothing\}$ acts as an order zero and $\mathcal{E}$ acts as an order unit for the partially ordered set $\mathcal{L}$.

For $[D] \in \mathcal{L}$, we can now form an absolute negation $[D]' = \{C \in \mathcal{O} \mid C \perp_m D\}$ as suggested above. It readily follows from the theorems stated in the previous section that the mapping $'\colon \mathcal{L} \to \mathcal{L}$ is a well-defined orthocomplementation on $\mathcal{L}$; that is, for $[D]$, $[C] \in \mathcal{L}$,

(i) $[D] \leqq [C]$ if and only if $[C]' \leqq [D]'$,
(ii) $([D'])' = [D]$ and
(iii) $[D] \wedge [D]' = \{\varnothing\}$, $[D] \vee [D]' = \mathcal{E}$.

Henceforth, as above, we shall use the symbols $\wedge$ and $\vee$ to denote infimum (greatest lower bound) and supremum (least upper bound) respectively in $\mathcal{L}$.

It is noteworthy that $\mathcal{L}$ is actually a complete lattice; that is, if $\{[D_\alpha] \mid \alpha \in I\}$ is any family of elements of $\mathcal{L}$, then the infimum $\wedge_\alpha [D_\alpha]$ and the supremum $\vee_\alpha [D_\alpha]$ exist and are given by the following convenient formulas:

(1) $\wedge_\alpha [D_\alpha] = [C]$, where C is *any* maximal orthogonal subset of $\bigcap_\alpha D_\alpha^{+}$.

(2) $$\vee_\alpha [D_\alpha] = (\wedge_\alpha [D_\alpha]')'.$$

It is significant that $\mathcal{L}$ is a distributive lattice (and, hence, a Boolean algebra) if and only if $\mathcal{P}$ contains exactly one primitive operation. However, a weak form of the distributive law, the so-called orthomodular identity, always obtains [**41**]:

(3) $$[C] \leqq [D] \Rightarrow [D] = [C] \vee ([D] \wedge [C]'), \qquad \text{for } [C], [D] \in \mathcal{L}.$$

Thus, $\mathcal{L}$ forms what is referred to in the literature as an *orthomodular lattice*—a type of lattice that has been extensively studied in recent years because of its connection with quantum theory and von Neumann algebras [7, 9, 10, 11, 13, 14, 16, 17, 18, 21, 22, 24, 28, 29, 30, 31, 33, 38, 41, 42, 43, 47].

Specifically, an orthomodular lattice is a lattice L with an order zero 0 and an order unit 1 and with an orthocomplementation $'\colon L \to L$ such that the orthomodular identity $e \leqq f \to f = e \vee (f \wedge e')$ holds for all $e, f \in L$. For $e, f \in L$, it is customary to say that e and f are *orthogonal*, and to write $e \perp f$, if and only if $e \leqq f'$. If every set of pairwise orthogonal elements of L is countable, then L is said to be *separable*. It is also common practice to say that e is compatible with f, and to write $e \leftrightarrow f$, if and only if there exist mutually orthogonal elements $e_1, f_1, d \in L$ such that $e = e_1 \vee d$ and $f = f_1 \vee d$. It is a fact that any maximal set of pairwise compatible elements of L is closed under the operations of $\wedge$, $\vee$ and $'$ and forms a Boolean algebra under these operations. These Boolean algebras are known as the *blocks* of L. Thus, the structure of L is determined by the structure of its blocks and the manner in which they "intertwine."

The intersection of all the blocks of an orthomodular lattice L is called the *center* of L and is denoted by $C(L)$. An element $e \in C(L)$ is called a *central element* and an element $e \in L$ is central if and only if $e \leftrightarrow f$ for every $f \in L$. One says that

L is *irreducible* if and only if 0 and 1 are the only central elements in L. An orthomodular lattice L is irreducible if and only if it cannot be factored (nontrivially) as a cartesian product. For the additional lattice theoretic terminology used below, see [**1**, **12**].

In $\mathcal{L}$, we already have seen that $[C]=[D]'$ if and only if there exist representative elements $C_1\in[C]$ and $D_1\in[D]$ such that $C_1\perp_m D_1$. Similarly, $[C]\leftrightarrow[D]$ and $[C]\perp[D]$ in $\mathcal{L}$ are respectively equivalent to $C_1\leftrightarrow D_1$ and $C_1\perp D_1$ in $\mathcal{O}$.

It can be shown that the calculus of events can be recaptured (up to an isomorphism) from the structure of $\mathcal{L}$ as an orthomodular lattice. As a consequence, a lattice automorphism of $\mathcal{L}$ automatically preserves orthocomplementation and is necessarily induced by a "change of variables"—permutations of outcomes in outcome sets and permutations of outcome sets of the same cardinality.

We shall say that an element $[D]\in\mathcal{L}$ is *bounded* if and only if there exists a bounded event $D_1\in\mathcal{O}$ such that $D_1\in[D]$. We denote by $\mathcal{L}_0$ the set of all bounded elements of $\mathcal{L}$. It is significant that $\mathcal{L}_0$ is a sublattice of $\mathcal{L}$ closed under the orthocomplementation (hence, $\mathcal{L}_0$ is an orthomodular lattice in its own right). The elements of $\mathcal{L}_0$ are in fact the experimental propositions that are in practice operationally accessible. We shall call $\mathcal{L}_0$ the *operational logic* over $\mathcal{P}$ and we shall call $\mathcal{L}$ the *complete operational logic* over $\mathcal{P}$.

THEOREM. *$\mathcal{L}$ is a complete orthomodular lattice, while $\mathcal{L}_0$ is an orthomodular lattice that is not σ-complete. The outcome set for every primitive operation $\omega\in\mathcal{P}$ is countable if and only if both $\mathcal{L}$ and $\mathcal{L}_0$ are separable. The orthomodular lattices $\mathcal{L}$ and $\mathcal{L}_0$ are simple (hence irreducible) if and only if $\mathcal{P}$ contains two or more primitive operations; otherwise, they are both Boolean algebras. Neither $\mathcal{L}$ nor $\mathcal{L}_0$ contains atoms, and consequently are order dense in themselves. The lattices $\mathcal{L}$ and $\mathcal{L}_0$ are modular if and only if they are Boolean algebras. Every element of $\mathcal{L}$ is a supremum of an orthogonal subset of $\mathcal{L}_0$; in fact, $\mathcal{L}$ is the completion by cuts of $\mathcal{L}_0$. If $\mathcal{P}$ contains two or more primitive operations, then every automorphism of $\mathcal{L}$ or of $\mathcal{L}_0$ is induced by a "change of variables."*

5. Conclusion. The above results only provide a glimpse of the exceedingly rich and descriptive structure of $\mathcal{L}$. We have not even considered the obvious question of the character of the sublogics that could be generated by various sets of operations.

Implicit in the structure of the logic $\mathcal{L}$, there is a complete description of the physical operations of concern and the relations between them. In the interest of brevity, however, we must postpone such details to later papers.

The so-called logic of quantum mechanics [**3**, **32**] is also a nonclassical orthomodular lattice; as a consequence, one might naively suppose that $\mathcal{L}$ would also be such a logic if the generating primitive operations were pertinent to quantum mechanical systems. Although this conjecture is demonstrably false, it nevertheless is so that there exist connections between $\mathcal{L}$ and quantum logic. The exact nature of these connections will be discussed in a subsequent paper. Since the

elements of $\mathcal{L}$ are explicitly defined in terms of physical operations, and this is not the case in quantum logic, it is important that these connections be understood.

It is noteworthy, for example, that in quantum logic the interpretation of the lattice theoretic infimum and supremum is a delicate matter [2, 3, 27], which many authors feel is still unresolved; whereas in $\mathcal{L}$ this issue can readily be settled. We need only consider the infimum, since the interpretation of $[D]'$ is transparent and $\bigvee_\alpha[D_\alpha] = (\bigwedge_\alpha[D_\alpha]')'$. Recall that $\bigwedge_\alpha[D_\alpha] = [C]$, where C is any maximal orthogonal subset of $\bigcap_\alpha D_\alpha^+$, and hence that $C^+ = \bigcap_\alpha D_\alpha^+$. Thus it is evident that the experimental proposition $\bigwedge_\alpha[D_\alpha]$ is validated by an outcome $b \in \Gamma$ if and only if b validates every $[D_\alpha]$. On the other hand, if an outcome $b \in \Gamma$ refutes any one of the experimental propositions $[D_\alpha]$, then it will refute $\bigwedge_\alpha[D_\alpha]$. This much would be expected; what might not be supposed is the fact that there exist, in general, $b \in \Gamma$ that refute $\bigwedge_\alpha[D_\alpha]$, but fail to refute any one of the experimental propositions $[D_\alpha]$. We also have that $\bigwedge_\alpha[D_\alpha] = \{\varnothing\}$ if and only if there does not exist an outcome $b \in \Gamma$ that will simultaneously validate every $[D_\alpha]$. Although this "interpretation" leaves something to be desired in the way of classical insight, it does specify precisely and operationally the conditions under which the experimental proposition $\bigwedge_\alpha[D_\alpha]$ is to be regarded as true or false—and this is all that is required.

The interpretation difficulty mentioned above should have come as no surprise, for, after all, the logic of quantum mechanics has been a consequence of a model of "reality" rather than a foundation for it. This state of affairs is by no means unique; in all of the empirical sciences, deterministic or stochastic models of nature are characteristically established first, and then—if at all—the implicit logics are investigated. In brief, these *posterior* logics depend not on reality, the purported ultimate authority of empirical science, but on models of reality. This curious practice, that leaves the conventional wisdom so vulnerable to crucial physical tests, such as the Michelson-Morley experiment, was noted by Mach [26] among others. Scientists [5, pp. 3–4], [46] have sought "self-healing" systems with which to escape the havoc caused by these periodic upheavals. It would seem that a large step in this direction will have been taken when we heed the advice of Garrett Birkhoff, offered in connection with quantum mechanics [2, p. 159]:

> "Scientifically, quantum logic should draw its authority directly from experiments. This approach is not only scientific; it has the mathematical advantage of making the lattice theory of the quantum logic autonomous."

The emphasis on the operational approach described above should not be construed as an adoption of radical empiricism. The development of a purely operational logic does not entail the simultaneous rejection of subjective methods; in particular, explicitly identifying the observables of an experimental science does not automatically deny the unifying and explanatory power of idealized models. The operational logics described here are only intended to

serve as foundations on which to erect the more familiar and utilitarian machinery of science. In subsequent papers a nonclassical statistics and an essentially subjective logic will be constructed on these foundations. Some of the details and a discussion of this program can be found in [39], [40].

This article is based in part on an address to the Northeastern Section of the MAA, November 30, 1968, by D. J. Foulis.

References

1. Garrett Birkhoff, Lattice Theory, Amer. Math. Soc. Colloquium Publ. XXV, 3rd ed., 1967.

2. ———, Lattices in applied mathematics, Proc. Symposia in Pure Math. II, Lattice Theory, Amer. Math. Soc., Providence, R.I. (1961), 155–184.

3. ———, and John von Neumann, The logic of quantum mechanics, Ann. of Math., 37 (1936) 823–843.

4. Niels Bohr, Essays 1958–1962 on Atomic Physics and Human Knowledge, Interscience, New York, 1963.

5. P. W. Bridgman, The Logic of Modern Physics, Macmillan, New York, 1927.

6. G. Bodiou, Théorie Dialectique des Probabilités, Gauthier-Villars, Paris, 1964.

7. D. E. Catlin, Irreducibility conditions on orthomodular lattices, J. Natural Sci. and Math., 8 (1968) 81–87.

8. ———, Spectral theory in quantum logics, Internat. J. Theor. Phys., 1 (1968) 285–297.

9. D. J. Foulis, Baer *-semigroups, Proc. Amer. Math. Soc., 11 (1960) 648–654.

10. ———, Conditions for the modularity of an orthomodular lattice, Pacific J. Math., 11 (1961) 889–895.

11. ———, A note on orthomodular lattices, Portugal. Math., 27 (1962) 65–72.

12. H. Gericke, Lattice Theory, Ungar, New York, 1966.

13. R. J. Greechie, On the structure of orthomodular lattices satisfying the chain condition, J. Comb. Theory, 4 (1968) 210–218.

14. ———, Hyper-irreducibility in an orthomodular lattice, J. Natural Sci. and Math., 8 (1968) 109–111.

15. S. P. Gudder, Spectral methods for a generalized probability theory, Trans. Amer. Math. Soc., 119 (1965) 428–442.

16. S. S. Holland, Jr., A Radon-Nikodym theorem in dimension lattices, Trans. Amer. Math. Soc., 108 (1963) 66–87.

17. ———, Distributivity and perspectivity in orthomodular lattices, Trans. Amer. Math. Soc., 112 (1964) 330–343.

18. ———, The Current Interest in Orthomodular Lattices, Trends in Lattice Theory, ed. J. C. Abbott, Van Nostrand, Princeton, N.J., 1969.

19. J. M. Jauch and C. Piron, Can hidden variables be excluded in quantum mechanics?, Helv. Phys. Acta, 36 (1963) 827–837.

20. J. M. Jauch, Foundations of Quantum Mechanics, Addison-Wesley, Reading, Mass., 1968.

21. M. F. Janowitz, Quantifiers and orthomodular lattices, Pacific J. Math., 13 (1963) 1241–1249.

22. ———, Quantifier theory on quasi-orthomodular lattices, Illinois J. Math., 9 (1965) 660–676.

23. A. N. Kolmogorov, Foundations of the Theory of Probability, 2nd ed., Chelsea, New York, 1956 (German ed. 1933).

24. L. H. Loomis, The Lattice Theoretic Background of the Dimension Theory of Operator Algebras, Mem. Amer. Math. Soc. Providence, R.I., 18 (1955).

25. G. Ludwig, Versuch einer axiomatischen Grundlegung der Quantenmechanik und allgemeinerer physikalischer Theorien, Z. Physik, 181 (1964) 233–260.

26. Ernst Mach, Die Mechanik in ihrer Entwicklung, 1st ed., 1883 (English transl., T. J. McCormack, Chicago, 1893).

27. M. D. MacLaren, Notes on Axioms for Quantum Mechanics, ANL 7065, Argonne National Lab., Argonne, Ill., 1965.

28. ———, Atomic orthocomplemented lattices, Pacific J. Math., 14 (1964) 597–612.

29. ———, Nearly modular orthocomplemented lattices, Trans. Amer. Math. Soc., 114 (1965) 401–416.

30. S. Maeda, Dimension functions on certain general lattices, J. Sci. Hiroshima Univ., Ser. A-1 Math., 19 (1955) 211–237.

31. ———, On conditions for the orthomodularity, Proc. Japan Acad., 42 (1966) 247–251.

32. G. W. Mackey, Mathematical Foundations of Quantum Mechanics, Benjamin, New York, 1963.

33. M. Nakamura, The permutability in a certain orthocomplemented lattice, Kodai Math. Sem. Rep., 9 (1957) 158–160.

34. J. C. T. Pool, Simultaneous Observability and the Logic of Quantum Mechanics, Thesis, State Univ. of Iowa, 1963.

35. ———, Semimodularity and the logic of quantum mechanics, Commun. Math. Phys., 9 (1968) 212–228.

36. ———, Baer *-semigroups and the logic of quantum mechanics, Commun. Math. Phys., 9 (1968) 118–141.

37. A. Ramsay, A theorem on two commuting observables, J. of Math., 15 (1966) 227–234.

38. ———, Dimension theory in complete orthcomplemented weakly modular lattices, Trans. Amer. Math. Soc., 116 (1965) 8–31.

39. C. H. Randall, A Mathematical Foundation for Empirical Science, Thesis, Rensselaer Polytechnic Institute, Troy, New York, 1966.

40. ———, A mathematical setting for inductive reasoning, SIAM Rev. (to be published).

41. C. H. Randall and D. J. Foulis, Lexicographic orthogonality, J. Combinatorial Theory (to be published).

42. U. Sasaki, On orthcomplemented lattices satisfying the exchange axiom, J. Sci. Hiroshima Univ., Ser. A-1 Math., 17 (1954) 293–302.

43. E. A. Schreiner, Modular pairs in orthomodular lattices, Pacific J. Math., 19 (1966) 519–528.

44. I. E. Segal, Postulates for general quantum mechanics, Ann. of Math., 48 (1947) 930–948.

45. ———, Mathematical Problems of Relativistic Physics, Amer. Math. Soc., 1963.

46. L. Tisza, The conceptual structure of physics, Rev. Mod. Phys., 35 (1963) 151–185.

47. D. Topping, Asymptoticity and semimodularity in projection lattices, Pacific J. Math., 20 (1967) 317–325.

48. V. S. Varadarajan, Geometry of Quantum Theory, vol. 1, Van Nostrand, Princeton, N.J., 1968.

49. J. von Neumann, Mathematical Foundations of Quantum Mechanics, Princeton Univ. Press, Princeton, N.J., 1955.

50. N. Zierler, Axioms for nonrelativistic quantum mechanics, Pacific J. Math., 11 (1961) 1151–1169.

HELLY TYPE THEOREMS DERIVED FROM BASIC SINGULAR HOMOLOGY

H. E. DEBRUNNER, University of Bern, Switzerland

In the first part of this paper the famous Helly intersection theorem is presented as a simple application of singular homology in a manner suitable for classroom use. In Section 6 the same proof is exploited to yield a partially new stronger result on intersections in n-manifolds.

1. Introduction. The tedious work needed to verify the basic properties of singular homology theory asks in introductory courses for a wealth of relevant applications. With restriction to the very basic facts and to simple spaces such as spheres and cells recent textbooks present as applications the invariance of dimension, the Brouwer fixed point theorem, the nonexistence of unit tangent vector fields on spheres of even dimension, the Jordan-Brouwer separation theorem, the invariance of domain and the fundamental theorem of algebra. As another application of the same restricted tools we shall derive the following Helly intersection theorem:

THEOREM 1. (Helly [5]). *Let $\{X_j\}_{j\in J}$ be a finite family of open convex subsets of euclidean n-space R^n such that each $n+1$ members of the family have a point in common. Then all the sets of the family have a point in common.*

This statement has numerous variants and applications; for an excellent survey on the ramifications and proofs of Helly's theorem the reader is referred to the Danzer-Grünbaum-Klee report [3]. This flexibility is related to the fact that the theorem has a purely topological background. In fact, Helly has also proved a more general but much less known topological version of this theorem. We shall present here this topological intersection theorem. The available proofs, Helly's in [6] and the simpler one in Alexandroff-Hopf [1], do not stand at the basis of homology theory as presented in introductory courses.

We shall use singular homology with integer coefficients. Any other nonzero coefficient group would do as well, for in the course of the proof we need only check whether some abelian groups are zero groups or not. The needed part of homology theory consists of the functorial properties, the fact that deformation retractions induce isomorphisms in homology, and the exactness of the Mayer-Vietoris sequence for two subspaces covering a space by their relative interiors; in addition we shall use the following weak continuity property:

(C). *If $\{U_\alpha\}$ is a family of subspaces of the topological space Y such that each compact subspace C of Y is contained in some space U_α of the family and such that $H_q(U_\alpha)=0$ for each space of the family, then $H_q(Y)=0$.*

For singular homology, (C) is clearly true, since the carrier of each singular chain c is a compact set C, and if a cycle c of Y is bounding in some U_α, the same holds in Y.

For simplicity we use for nonempty spaces Y also the reduced homology groups $\tilde{H}_q(Y)$, thus saving extra considerations for $q=0$.

2. Statements. Let a *homology cell* be a space having singular homology groups isomorphic to those of a point or cell, i.e., a homology cell X is a *nonempty* space with $\tilde{H}_q(X)=0$ for all integers q. Since nonempty convex subsets of R^n are homology cells and intersect in convex subsets (eventually empty), Theorem 1 is contained in

THEOREM 2. (Helly [6]). *Let $\{X_j\}_{j\in J}$ be a finite family of open subsets of euclidean n-space R^n such that the intersection $X_{j_1}\cap \cdots \cap X_{j_r}$ of each r sets of this family is nonempty for $r\leqq n+1$ and is even a homology cell for $r\leqq n$. Then $\bigcap_{j\in J} X_j$ is a homology cell.*

The proof of Theorem 2 is based on two lemmata. The first tells us something about the homology structure of spaces built up from open homology cells.

LEMMA A_m. *Let $\{X_1, \cdots, X_m\}$ be a finite family of homology cells imbedded as open subsets in some topological space M in such a way that the intersection of each r members of the family is a homology cell whenever $r=1, \cdots, m-1$.*

(a) *If $X_1\cap \cdots \cap X_m$ is empty, then $\tilde{H}_q(X_1\cup \cdots \cup X_m)$ is nonzero for $q=m-2$ and zero for $q\neq m-2$.*

(b) *If $X_1\cap \cdots \cap X_m$ is nonempty, then $\tilde{H}_q(X_1\cup \cdots \cup X_m)\approx \tilde{H}_{q-m+1}(X_1\cap \cdots \cap X_m)$ for each integer q. In particular $X_1\cup \cdots \cup X_m$ is a homology cell if and only if $X_1\cap \cdots \cap X_m$ is a homology cell.*

The second lemma exhibits a significant property of dimension.

LEMMA B_n. *If Y is a nonempty open subset of R^n, then $\tilde{H}_q(Y)=0$ for $q\geqq n$.*

Note that the corresponding statement for closed subsets of R^n is not valid in singular homology theory, as shown in [2].

3. Proof of Theorem 2. Assume there exist families satisfying the hypotheses of Theorem 2 but not the conclusion. Let $\{X_1, \cdots, X_m\}$ among such families be one with the minimal number m of members. We may apply Theorem 2 to any proper subfamily of $\{X_1, \cdots, X_m\}$, for the hypotheses pass to subfamilies, and due to the minimality of m we conclude that the intersection of all sets of any proper subfamily is a homology cell. This means that Lemma A_m applies to the chosen family $\{X_1, \cdots, X_m\}$. Consider first the case $X_1\cap \cdots \cap X_m=\emptyset$, which under the assumptions of Theorem 2 can happen only if $m>n+1$. Lemma A_m gives the result (i) $\tilde{H}_{m-2}(X_1\cup \cdots \cup X_m)\neq 0$. On the other hand, since $m-2\geqq n$, Lemma B_n shows (ii) $\tilde{H}_{m-2}(X_1\cup \cdots \cup X_m)=0$. The statements (i) and (ii) contradict. Consider now the case $X_1\cap \cdots \cap X_m\neq\emptyset$. Our choice of the family was such that $X_1\cap \cdots \cap X_m$ is not a homology cell, and under the hypotheses of Theorem 2 we thus must have $m>n$. There exists $p\geqq 0$ such that $\tilde{H}_p(X_1\cap \cdots \cap X_m)\neq 0$ and application of Lemma A_m (b) gives (i) $\tilde{H}_{p+m-1}(X_1\cup \cdots \cup X_m)\neq 0$. On the other hand, $m>n$

guarantees $p+m-1 \geqq n$, so Lemma B_n yields (ii) $\tilde{H}_{p+m-1}(X_1 \cup \cdots \cup X_m)=0$. Again (i) and (ii) contradict, and the proof is reduced to the lemmata.

4. Proof of Lemma A_m. We proceed by induction on m. Suppose $m>2$ and the statement A_{m-1} is valid. Let $\{X_1, \cdots, X_m\}$ be a family satisfying the common hypothesis of A_m. The subfamily $\{X_1, \cdots, X_{m-1}\}$ then clearly satisfies the assumptions of A_{m-1} and the particular case of A_{m-1} (b) guarantees that $X_1 \cup \cdots \cup X_{m-1}$ is a homology cell U_1. There is a reduced Mayer-Vietoris sequence for the pair $\{X_1 \cup \cdots \cup X_{m-1}, X_m\}$ of open homology cells, namely

$$\cdots \to \tilde{H}_q(U_1) \oplus \tilde{H}_q(U_2) \to \tilde{H}_q(U_1 \cup U_2) \xrightarrow{\partial_*} \tilde{H}_{q-1}(U_1 \cap U_2) \to \tilde{H}_{q-1}(U_1) \oplus \tilde{H}_{q-1}(U_2) \to \cdots \tag{1}$$

where $U_1 = X_1 \cup \cdots \cup X_{m-1}$, $U_2 = X_m$, $U_1 \cap U_2 = (X_1 \cap X_m) \cup \cdots \cup (X_{m-1} \cap X_m)$. We know that U_1 and U_2 are homology cells, so the direct sum terms are zero and by exactness we have $\tilde{H}_q(U_1 \cup U_2) \approx \tilde{H}_{q-1}(U_1 \cap U_2)$, or explicitly

$$\partial_*: \tilde{H}_q(X_1 \cup \cdots \cup X_m) \approx \tilde{H}_{q-1}((X_1 \cap X_m) \cup \cdots \cup (X_{m-1} \cap X_m)). \tag{2}$$

Now apply A_{m-1} to the family $\{X_1 \cap X_m, \cdots, X_{m-1} \cap X_m\}$; note that intersections of r members in this family are intersections of $r+1$ members in the given family $\{X_1, \cdots, X_m\}$; in particular the intersection of all the members is equal to $X_1 \cap \cdots \cap X_m$. If this is empty, A_{m-1} (a) shows that the right hand side of (2) is nonzero for $q-1=(m-1)-2$ and zero otherwise, i.e., for $q=m-2$ and $q \neq m-2$ respectively. If $X_1 \cap \cdots \cap X_m$ is nonempty, the right hand side of (2) is by A_{m-1} (b) isomorphic to $\tilde{H}_{(q-1)-(m-1)+1}(X_1 \cap \cdots \cap X_m)$, so

$$\tilde{H}_q(X_1 \cup \cdots \cup X_m) \approx \tilde{H}_{q-m+1}(X_1 \cap \cdots \cap X_m).$$

Thus A_m is valid if A_{m-1} is.

It remains to start the induction with the case $m=2$. Let X_1, X_2 be open homology cells in M. If $X_1 \cap X_2$ is empty as in case A_2(a) then $X_1 \cup X_2$ is the disjoint union of two homology cells X_1, X_2, thus $\tilde{H}_0(X_1 \cup X_2) \neq 0$, $\tilde{H}_q(X_1 \cup X_2)=0$ for $q \neq 0$, as asserted.

If $X_1 \cap X_2$ is nonempty, the sequence (1) with the homology cells X_i in place of U_i again shows that $\partial_*: \tilde{H}_q(X_1 \cup X_2) \approx \tilde{H}_{q-1}(X_1 \cap X_2)$, which is statement A_2(b).

5. Proof of Lemma B_n. The statement B_n being trivial for $n=0$, we suppose $n \geqq 1$. Divide R^n by $(n-1)$-planes into a tessellation by congruent cubes and define S to be the family of all these closed nonoverlapping n-cubes filling up R^n. A closed subset of R^n which is the union of k n-cubes of S is called n-cubical of order k. To each open subset Y of R^n and each compact $C \subset Y$ there is a sufficiently fine such subdivision of R^n and with respect to it an n-cubical set U of finite order satisfying $C \subset U \subset Y$. In view of the continuity property (C) it suffices to verify the statement

$D(n, k)$: *If $U \subset R^n$ is n-cubical of order $k \geqq 1$, then $\tilde{H}_q(U)=0$ for $q \geqq n \geqq 1$.*

Of course this is trivial if the isomorphism between singular and simplicial

homology is available. Avoiding it, we prove $D(n, k)$ by simultaneous induction on n and k. The start of the induction is trivial: In the case $D(1, k)$, U is a finite union of disjoint intervals, in the case $D(n, 1)$, U is a cube, thus both times $\tilde{H}_q(U)=0$ for $q \geqq n \geqq 1$. Now we consider with respect to a tessellation S an n-cubical set U of order k, where $n>1$ and $k>1$, and we assume $D(n', k')$ is valid if $n'<n$ or $k'<k$. If U is not pathconnected, it is the union of two disjoint n-cubical sets U_1, U_2 of order $<k$ and

$$\tilde{H}_q(U) \approx \tilde{H}_q(U_1) \oplus \tilde{H}_q(U_2) = 0 \qquad \text{for } q \geqq n \geqq 1$$

by the inductive hypothesis applied to U_1 and U_2, and we are done. So assume U is pathconnected. Let E_t denote the $(n-1)$-plane given by $x_n=t$ in a cartesian coordinate system with axes and unit such that the cube

$$\{x \in R^n \mid 0 \leqq x_i \leqq 1 \qquad \text{for } i = 1, \cdots, n\}$$

is a cube of our tessellation S. Define t_1 and t_2 to be the infimum and the supremum, respectively, of $\{t \in R \mid E_t \cap U \neq \varnothing\}$. Pick out E_t for $t=\frac{1}{2}(t_1+t_2)$ and write E_0 for this $(n-1)$-plane and H_1, H_2 for the two closed half-spaces determined by E_0.

If U_0, U_1, U_2 respectively denote the union of all those n-cubes of S which are contained in U and meet, respectively, the plane E_0, the halfspace H_1, the halfspace H_2, then obviously $U=U_1 \cup U_2$ and $U_0=U_1 \cap U_2 \neq \varnothing$ (since U is connected). In the case $|t_2-t_1| \leqq 2$ there are at most two layers of cubes and we have even $U=U_1=U_2=U_0$. But the orthogonal projection of U_0 onto the $(n-1)$-cubical set $U_0 \cap E_0$ is a deformation retraction, so (i) $\tilde{H}_q(U_1 \cap U_2)=\tilde{H}_q(U_0) \approx \tilde{H}_q(U_0 \cap E_0)=0$ for $q \geqq n-1 \geqq 1$ by induction hypothesis. In the case $|t_2-t_1| \leqq 2$ we are done, since $U=U_0$. In the case $|t_1-t_2|>2$ we insert (i) in the Mayer-Vietoris sequence (1) set up for our sets U_1 and U_2 and by exactness we get (ii) $\tilde{H}_q(U)=\tilde{H}_q(U_1 \cup U_2) \approx \tilde{H}_q(U_1) \oplus \tilde{H}_q(U_2)$ for $q \geqq n$; but both U_1 and U_2 are n-cubical of order $<k$ and so the inductive hypothesis and (ii) yield $\tilde{H}_q(U)=0$ for $q \geqq n$. This completes the proof.

6. Generalizations. Helly's Theorem 1 on convex sets has a well-known analogue for convex subsets of the metrical n-sphere S^n, but in this variant the critical number is $n+2$ rather than $n+1$. In fact, in all the statements and proofs of Sections 2 and 3 we may obviously substitute S^{n-1} in place of R^n and a shift of the dimension number yields our assertion. Here a subset Y of the metrical sphere S is said to be *convex* if between any two points of Y there exists in S exactly one shortest geodesic arc and it is entirely contained in Y. The same definition applies to sets Y in any Riemannian manifold S; we shall see that *Theorem* 1 *also holds with R^n replaced by any Riemannian n-manifold S; only in the case S has the homology of an n-sphere an additional condition has to be imposed on the family or the critical number will be $n+2$ rather than $n+1$.* In this sense homology-n-spheres (i.e., n-manifolds having the homology of an n-sphere) are the only manifolds playing an exceptional role in the context of the Helly Intersection Theorem; by an additional argument involving the fundamental group it would

be possible to restrict the exceptions to homotopy spheres. These statements on intersections of convex sets again stem from a topological intersection theorem which we shall now deduce by combining a full exploitation of Lemma A_m with the analogue of Lemma B_n for n-manifolds.

Manifolds will be without boundary and connected. A space X is said to be p-acyclic, if it is nonempty and if $\tilde{H}_q(X)=0$ for all $q\leqq p$; in particular X is (-1)-acyclic if it is nonempty, and it is p-acyclic for all integers p if and only if it is a homology cell.

LEMMA B_n'. $(n\geqq 1)$. *If Y is an n-manifold, then $\tilde{H}_q(Y)=0$ for $q>n$; furthermore $\tilde{H}_n(Y)\approx Z$ or $\tilde{H}_n(Y)=0$ according to whether Y is compact and orientable or not.*

For a proof cf. [**4**, p. 121]. We now state our result, which contains theorems of Molnar [**7, 8, 9**] and of Berstein (see [**3**, p. 125]) as special cases.

THEOREM 3. *Let M denote an n-manifold which is not a homology-n-sphere and let $\{X_j\}_{j\in J}$ be a finite family of open subsets of M such that, for each choice of r members of this family, whenever $r\leqq n+1$ the intersection $X_{j_1}\cap\cdots\cap X_{j_r}$ is $(n-r)$-acyclic. Then $\bigcap_{j\in J}X_j$ is a homology cell. If M is a homology-n-sphere the same statement holds under the additional condition that no $n+2$ sets of the family cover M.*

Proof. Assume there exist families in M satisfying the hypotheses of Theorem 3 but not the conclusion. Let $\{X_1,\cdots,X_m\}$ among such families be one with the minimal number m of members. Theorem 3 applies to any proper subfamily of $\{X_1,\cdots,X_m\}$, for the hypotheses pass to subfamilies, and due to the minimality of m we conclude that the intersection of the sets of any proper subfamily of $\{X_1,\cdots,X_m\}$ is a homology cell. This means that Lemma A_m applies to the chosen family.

Assume $X_1\cap\cdots\cap X_m=\varnothing$. Under the assumptions of Theorem 3 this happens only if (i) $m\geqq n+2$. Write $Y=X_1\cup\cdots\cup X_m$. Lemma A_m(a) assures (ii) $\tilde{H}_{m-2}(Y)\neq 0$ and (iii) $\tilde{H}_q(Y)=0$ for $q\neq m-2$. But Y is itself an n-submanifold of M, so Lemma B_n' contrasted to (ii) yields $m-2\leqq n$ and comparison with (i) shows $m-2=n$. But then (ii) together with Lemma B_n' yield that $Y=M$, and by (ii) and (iii) M must be a homology-n-sphere. Yet in this case, since $m=n+2$, the additional condition imposed on $\{X_1,\cdots,X_m\}$ in Theorem 3 assures $Y\neq M$ and we arrive at a contradiction.

Now assume $X_1\cap\cdots\cap X_m\neq\varnothing$. By the choice of our family $X_1\cap\cdots\cap X_m$ is not a homology cell, so there exists a smallest integer q such that (iv) $\tilde{H}_{q-m+1}(X_1\cap\cdots\cap X_m)\neq 0$ and consequently, by Lemma A_m, also (v) $\tilde{H}_q(X_1\cup\cdots\cup X_m)\neq 0$. Since $X_1\cup\cdots\cup X_m$ is an n-manifold it follows $0\leqq q-m+1\leqq q\leqq n$ or equivalently (vi) $m\leqq q+1\leqq n+1$. Therefore, on the assumptions of Theorem 3, $X_1\cap\cdots\cap X_m$ is $(n-m)$-acyclic, hence $q-m+1>n-m$ by (iv). This and (vi) imply that $q=n$. But then (v) and the minimality of q show that $X_1\cup\cdots\cup X_m$ has the homology of an n-sphere and therefore

M is a homology-n-sphere covered by the sets $X_1, \cdots, X_m$. Since $m \leqq n+1$ again we arrive at a contradiction to the additional condition that M is not covered by less than $n+3$ sets of the given family.

References

1. P. Alexandroff and H. Hopf, Topologie, Springer, Berlin, 1935.
2. M. G. Barratt and J. Milnor, An example of anomalous singular homology, Proc. Amer. Math. Soc., 13 (1962) 293–297.
3. L. Danzer, B. Grünbaum and V. Klee, Helly's theorem and its relatives, Proc. of Symposia in Pure Mathematics, Amer. Math. Soc., 7 (1963). Convexity, 100–181.
4. M. Greenberg, Lectures on Algebraic Topology, Benjamin, New York, 1967.
5. E. Helly, Über Mengen konvexer Körper mit gemeinschaftlichen Punkten, Jber. Deutsch. Math. Verein., 32 (1923) 175–176.
6. ———, Über Systeme von abgeschlossenen Mengen mit gemeinschaftlichen Punkten, Monatsh. Math. und Physik, 37 (1930) 281–302.
7. J. Molnár, Über eine Verallgemeinerung auf die Kugelfläche eines topologischen Satzes von Helly, Acta Math. (Budapest), 7 (1956) 107–108.
8. ———, Über eine Vermutung von G. Hajós, *ibid.*, 8 (1957) 311–314.
9. ———, Über eine Übertragung des Hellyschen Satzes in sphärische Räume, *ibid.*, 8 (1957) 315–318.

Reader's Comments on Some Papers in this MONTHLY

R. A. ROSENBAUM, Wesleyan University

A. F. Horadam, *A Generalized Fibonacci Sequence*, 68 (1961) 455–460.

Miss Sue Dunfee of Hollins College points out that equations (13) and (17) should read as follows:

$$H_n^3 + H_{n+1}^3 = 2H_nH_{n+1}^2 + (-1)^n ef,$$

where $f=q$ for $n=1$, and $f=H_{n-1}$ for $n>1$;

$$\frac{H_{n+r} + (-1)^r H_{n-r}}{H_n} = a^r + b^r.$$

C. F. Harding, *Matrix of Finite Groups*, 74 (1967) 430–432.

Professor J. L. Berggren of Simon Fraser University remarks that, on pp. 32–34 of "Representation Theory of Finite Groups and Associative Algebras" by C. W. Curtis and Irving Reiner, the authors describe a method by which the regular matrix representation can be read off from a suitably arranged group table, thus contradicting a statement in the article.

M. L. Tikoo, *Location of the Zeroes of a Polynomial*, 74 (1967) 688–690.

Professor Marcia Ascher of Ithaca College notes that the corollary of Theorem 1 should contain the inequality

$$|Z| \leqq \max\left(|C_{i-1}|/|C_i|\,t,\ |C_{n-2}|\,t^{n-1},\ \sum_{i=1}^{n-1}(1/|C_n|\,t^i) + (|C_{n-1}|/|C_n|)\right),$$

$i=1, \cdots, n-2$, where $t>0$ is arbitrary, rather than the stated bound on $|Z|$, the latter being correct only for a certain class of t's.

J. C. Parnami, *On Iterates of Euler's ϕ-Function*, 74 (1967) 967–968.

Professor George K. White of the University of British Columbia and Professor M. G. Beumer of the Technological University of Delft call attention to some prior work:

Pillai, S. S., "On a function connected with $\phi(n)$," BAMS, 35 (1929) 837–841.

Shapiro, H. N., "On an arithmetic function arising from the ϕ-function," this MONTHLY, 50 (1943) 18–30.

A. A. Gioia, and A. M. Vaidya, *Amicable Numbers with Opposite Parity*, 74 (1967) 969–973.

Professor Thomas E. Elsner of the General Motors Institute writes that the last of the inequalities preceding (2.1) should read:

$$2^{r+1}\prod 2^{2\alpha_i a_i+1} > 2^{2b-1};$$

fortunately, this change does not alter the results of the paper.

MATHEMATICAL NOTES

EDITED BY DAVID DRASIN

Manuscripts for this Department should be sent to David Draisin, Division of Mathematical Sciences, Purdue University, Lafayette, IN 47907.

AN APPLICATION OF SIMPLICIAL HOMOLOGY THEORY

GAIL H. ATNEOSEN, Western Washington State College

The purpose of this note is to illustrate an elementary application of simplicial homology theory to obtain a geometric result.

THEOREM. *Every polygonal knot in Euclidean 3-dimensional space E^3 spans a surface in E^3.*

A *surface* is a compact connected 2-manifold; a knot *spans* a surface if it is the manifold boundary of the surface. The result is not new. It is well known [**6**, p. 572] that every oriented polygonal knot in E^3 spans an orientable surface in E^3 (see also [**1**, p. 140] and [**3**]). In fact, J. Levine has given a proof of the generalization to locally flat n-spheres in R^{n+2} for $n \geqq 2$ using transversality [**5**, p. 11]. He has further used this to give a classification of knots in codimension 2 for $n \geqq 4$; generalizations to $n=3$ may be obtained using results of J. L. Shaneson [**7**].

The proof we give of the above theorem could be presented immediately after the development of the simplicial homology groups. This would provide the student with an opportunity to work with the newly acquired notions in a

strongly geometric setting. Some details of the proofs are omitted, but they can be readily supplied.

First, we observe that all polygonal knots in E^3 can be carried by a space homeomorphism into a polygonal knot that lies in a 3-book. A *3-book* is a subset of E^3 that is the union of three geometric 2-simplexes such that each pair of 2-simplexes meets precisely on a single geometric 1-simplex B on the face of each. The 2-simplexes are called the *leaves* of the 3-book and B is its *back*. A polygonal knot K is said to be *nicely embedded* in a 3-book if: (i) K intersected with the back of the 3-book does not contain an arc; (ii) if p is a point of K contained in the back of the 3-book and s and t are straight line segments of K with endpoint p, then s and t are contained in distinct leaves of the 3-book. A figure-eight knot that is nicely embedded in a 3-book is pictured in Figure 1.

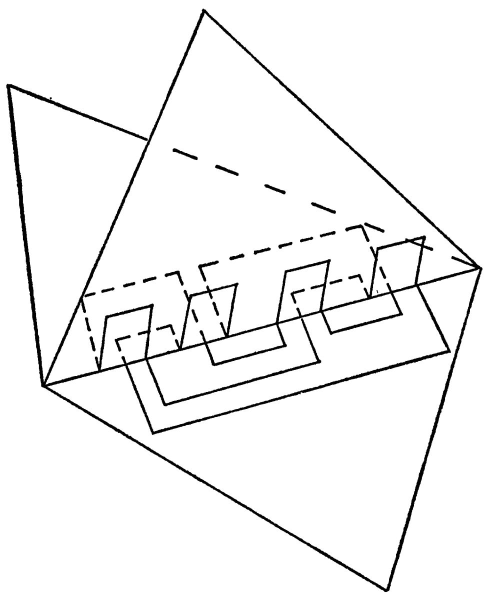


FIG. 1

LEMMA 1. *Let K be a polygonal knot in E^3. Then there is a space homeomorphism mapping E^3 onto E^3 such that the image of K is a polygonal knot that is nicely embedded in a 3-book.*

Proof. We may assume that K is in regular position [2, p. 6] and that all the double points of a parallel projection of K onto the xy plane are contained in the x-axis. There is a space homeomorphism which carries the over-crossings into the upper half of the xz plane and is such that the entire image of K is contained in the union of the xy plane and the upper half of the xz plane. It then follows that there is a 3-book containing the image of K. A slight modification of the above homeomorphism, if necessary, will yield a nice embedding of the image of K in a 3-book.

LEMMA 2. *Let K be a polygonal knot nicely embedded in a 3-book. Then K spans a surface in the 3-book.*

Proof. We may assume that the 3-book B^3 is triangulated in such a manner that K is the carrier of a subcomplex of the triangulation. This triangulation will be fixed in the following discussion, and it will be used in computing all simplicial homology groups. For the coefficient group we shall use the group Z_2 of integers modulo two. Then $H_1(B^3, Z_2)$, the first simplicial homology group of B^3 with coefficients in Z_2, is trivial. Thus, every 1-cycle modulo two is also a boundary. In particular, if c_1 is the 1-cycle modulo two that has value one on each 1-simplex in K and value zero on all other 1-simplexes, then there exists a 2-chain modulo two c_2 such that $\partial c_2 = c_1$ (∂ denotes the boundary operator). Let M be the subset of E^3 that is the union of all 2-simplexes in the triangulation of B^3 on which c_2 has nonzero value.

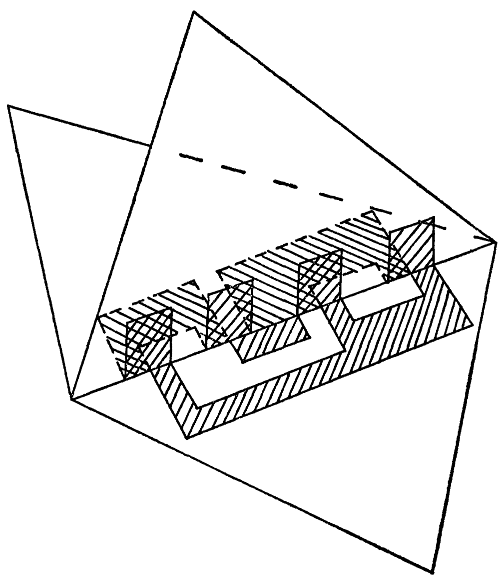


FIG. 2

We next prove that M is a surface with manifold boundary K. Since $\partial c_2 = c_1$, every 1-simplex in $M-K$ must be the face of precisely two 2-simplexes in M. Thus if $x \in M-K$, then x has a neighborhood in M homeomorphic to an open 2-cell. If $x \in K$ and x is not a vertex in the triangulation of B^3, then, since K is nicely embedded and $\partial c_2 = c_1$, it follows that x is contained in the interior of a 1-simplex that is the face of exactly one 2-simplex in M. If $x \in K$ and x is a vertex, then straightforward arguments considering various cases yield that x has a neighborhood in M whose closure in M is homeomorphic to a closed 2-cell. Thus M is a surface in B^3 with manifold boundary K.

The theorem then follows from the above two lemmas. Figure 2 shows a

surface spanned by a figure-eight knot that is nicely embedded in a 3-book. This surface is nonorientable.

References

1. R. H. Fox, A Quick Trip Through Knot Theory, Topology of 3-Manifolds and Related Topics, Prentice-Hall, Englewood Cliffs, N. J., 1962.

2. ——— and R. M. Crowell, Introduction to Knot Theory, Ginn, Boston, Mass., 1963.

3. R. E. Goodrick, A note on Seifert circles, Proc. Amer. Math. Soc., 21 (1969) 615–617.

4. J. G. Hocking and G. S. Young, Topology, Addison-Wesley, Reading, Mass., 1961.

5. J. Levine, Unknotting spheres in codimension two, Topology, 4 (1965) 9–16.

6. H. Seifert, Über das Geschlecht von Knoten, Math. Ann., 110 (1934) 571–592.

7. J. L. Shaneson, Embeddings with codimension two of spheres in spheres and h-cobordisms of $S^1 \times S^3$, Bull. Amer. Math. Soc., 74 (1968) 972–974.

ON LOCALLY RECURRENT FUNCTIONS

TIBOR ŠALÁT, University of Komenský, Czechoslovakia

A real function f defined on an interval $[a, b]$ is said to be *locally recurrent* on this interval if for every $x \in [a, b]$ and every $\epsilon > 0$ there exists y with $f(y) = f(x)$ and $0 < |y - x| < \epsilon$ (cf. [1]).

THEOREM. *Let f be a nonconstant continuous locally recurrent function on $[a, b]$. Then f is a function of unbounded variation on $[a, b]$.*

This has been proved by Marcus [3] but we establish it as a consequence of a theorem on the Banach indicatrix.

Proof. Let τ denote the Banach indicatrix of f on $[a, b]$. Thus if y is a real number let $\tau(y)$ be the number of elements of the set

$$\{x \in [a, b]; f(x) = y\} \tag{1}$$

if this set is finite and $\tau(y) = \infty$ if the set (1) is infinite. We also denote by $V_a^b(f)$ the variation of f on $[a, b]$.

Since f is a nonconstant continuous function on $[a, b]$ we have

$$m = \min_{x \in [a,b]} f(x) < \max_{x \in [a,b]} f(x) = M.$$

It follows from the well-known Darbouxian property of the continuous functions that for every $y \in [m, M]$ there exists an $x \in [a, b]$ such that $f(x) = y$. So we have $\tau(y) \geqq 1$ for every $y \in [m, M]$. From this and from the fact that f is a locally recurrent function it is easy to verify that

$$\tau(y) = + \infty \tag{2}$$

for every $y \in [m, M]$.

We use the relation

$$V_a^b(f) = \int_{-\infty}^{+\infty} \tau(y)dy = \int_m^M \tau(y)dy$$

(cf. [5] p. 374–375 or [6] p. 246) with (2) and the theorem follows at once.

References

1. K. A. Bush, Locally recurrent functions, this MONTHLY, 69 (1962) 199–206.
2. J. G. Mauldon, The differentiability of locally recurrent functions, this MONTHLY, 72 (1965) 983–985.
3. S. Marcus, Some remarks on real functions II, Rev. Math. Pures Appl., 11 (1966) 911–916.
4. ———, On locally recurrent functions, this MONTHLY, 70 (1963) 822–826.
5. R. Sikorski, Funkcje rzeczywiste I, Warszawa, 1958.
6. I. P. Natanson, Theory of functions of real variable, (Russian) Moskva, 1957.

HAMEL VERSUS SCHAUDER DIMENSION

J. W. EVANS, University of California, and R. A. TAPIA, University of Wisconsin

In this note we define the notion of Schauder dimension of a Banach space and show that the Schauder dimension is less than the Hamel dimension, as is known to be true for separable Banach spaces, if and only if the Schauder dimension possesses a property which the cardinal $\aleph_0$ possesses. We shall denote the cardinality of a set A by $|A|$. Let X be a real Banach space and M a subset of X.

DEFINITION 1. *The set M is weakly (strongly) linearly independent if it is not contained in the linear (closed linear) span of any proper subset.*

DEFINITION 2. *A maximal weakly (strongly) linearly independent subset of X is called a Hamel (extended Schauder) basis for X.*

DEFINITION 3. *The cardinality of any Hamel (extended Schauder) basis is called the Hamel (Schauder) dimension of X.*

REMARK. A set which is strongly linearly independent is often called *topologically free*, while extended Schauder bases are often referred to as *minimal generating sets*. It is extremely difficult to extend the concept of a Schauder basis from separable Banach spaces to arbitrary Banach spaces if every element in the space is still required to have a unique representation in terms of the basis, as is true in the separable case. This difficulty arises from the subtleties involved in attempting to formulate a workable definition of conditional convergence of series summed over an arbitrary countable set. The obvious generalization of merely introducing a linear ordering or well-ordering on the index set, as is suggested in [1] and [7], will not work. Markouchevitch [5] very cleverly drops this requirement and gives the following generalization of a Schauder basis.

DEFINITION 4. *The set $\{x_\alpha | \alpha \in A\} \subset X$ is called a Markouchevitch basis for X if*

(i) *(biorthogonality) there exists a family of continuous linear functionals $\{f_\alpha | \alpha \in A\}$ such that $f_\alpha(x_\beta) = \delta_{\alpha\beta}$ (Kronecker delta);*

(1) (ii) *(totality of x_α) for any continuous linear functional f, if $f(x_\alpha) = 0$ for all α then $f = 0$; and*

(iii) *(totality of f_α) for any $x \in X$, if $f_\alpha(x) = 0$ for all α then $x = 0$.*

REMARK. It is clear that any Schauder basis and (in Hilbert space) any complete orthonormal set satisfy these conditions. In fact, in Hilbert space conditions (ii) and (iii) are equivalent. The following proposition shows that a Markouchevitch basis is an extended Schauder basis.

PROPOSITION 1. *The set* $\{x_\alpha | \alpha \in A\} \subset X$ *is an extended Schauder basis for* X *if and only if conditions* (i) *and* (ii) *of* (1) *are satisfied.*

The proof is an immediate application of the Hahn-Banach theorem [3, p. 62].

REMARK. It is interesting to note that if continuity is not required in conditions (i) and (ii) of (1), then Proposition 1 holds with extended Schauder replaced by Hamel. Our definition of extended Schauder basis amounts to dropping requirement (iii) in (1).

Some authors follow Arsove and Edwards [1] and drop requirement (ii) from (1). In doing this certain paradoxes may arise, e.g., nonseparable spaces may have countable bases, see [1]. It seems very unnatural to allow bases whose linear span is not dense in the space, i.e., do not satisfy (ii) of (1).

PROPOSITION 2. *Any two Hamel* (*extended Schauder*) *bases for* X *have the same cardinalities.*

Proof. Let $\{x_\alpha | \alpha \in A\}$ and $\{y_\beta | \beta \in B\}$ be two Hamel (extended Schauder) bases for X. Assume $|A| < |B|$. To each x_α corresponds a finite (countable) set $B_\alpha \subset B$ such that x_α is contained in the linear (closed linear) span of $\{y_\beta | \beta \in B_\alpha\}$. Let $D = \bigcup_{\alpha \in A} B_\alpha$. We have $|D| \leqq \aleph_0 \cdot |A| = |A| < |B|$, therefore $\{y_\beta | \beta \in D\}$ is a proper subset of $\{y_\beta | \beta \in B\}$ whose linear (closed linear) span is all of X. We arrive at a similar contradiction by assuming $|A| > |B|$.

PROPOSITION 3. *If* B *is any cardinal, then there exist Banach spaces with Schauder dimension* B.

Proof. Let Ω be any point set with $|\Omega| = B$. It is well known that $L_p(\Omega)$ ($p \geqq 1$), with the discrete measure in Ω, is a Banach space, and it follows that $\{I_\omega | \omega \in \Omega\}$, where $I_\alpha(\beta) = \delta_{\alpha\beta}$, $\alpha, \beta \in \Omega$, is an extended Schauder basis for $L_p(\Omega)$.

DEFINITION 5. *A cardinal is of type I if it is the cardinality of a set which is the union of an infinite sequence of sets with strictly increasing cardinalities.*

THEOREM 1. *If* H *is a Hamel basis and* S *an extended Schauder basis for* X, *then*

(i) $|H| = 2^{|S|}$ *if* $|S|$ *is of type I and*

(ii) $|H| = |S|$ *if* $|S|$ *is not of type I.*

Proof. If $|S|$ is finite, then $|S|$ is clearly not of type I and the two notions of bases coincide. When $|S| \geqq \aleph_0$, we know from [3, p. 56 and 75] that $|X| = |H|$ and that the number of finite subsets of S is $|S|$, while the number

of countable subsets of S is $|S|^{\aleph_0}$. Each $x\in X$ is the limit of a sequence with terms in the linear span of S. There are at most $|S|^{\aleph_0}$ such sequences so $|X|\leqq|S|^{\aleph_0}$; but using (i) of (1) we see that to each distinct countable subset $\{x_i|i=1, 2, \cdots\}\subset S$ we may associate distinct

$$x=\sum\frac{x_i}{2^i\|x_i\|}\in X,$$

so that $|S|^{\aleph_0}\leqq X$ or $X=|S|^{\aleph_0}$. This gives us $|S|\leqq|S|^{\aleph_0}=|X|=|H|\leqq 2^{|S|}$. The theorem now follows, upon accepting the generalized continuum hypothesis, from the fact that $|S|<|S|^{\aleph_0}$ if and only if $|S|$ is of type I [6].

COROLLARY 1. *There exist real Banach spaces of equal Hamel dimension, but unequal Schauder dimension.*

Proof. If B is any cardinal of type I, then by Proposition 3 there exist Banach spaces with Schauder dimension B and Schauder dimension 2^B, but from the theorem both these spaces have Hamel dimension 2^B.

COROLLARY 2 (Goffman). *If B is any cardinal, then there exist Banach spaces with Hamel dimension B if and only if B is not of type I.*

Proof. If B is not of type I, then by the theorem the Banach space in Proposition 3 also has Hamel dimension B. Now let X be a real Banach space with Hamel basis

$$\{x_\alpha\mid\alpha\in D\}\text{ and }\quad D=\bigcup_{i=1}^{\aleph_0}A_i,\quad\text{where}\quad |A_{i+1}|>|A_i|.$$

If D_n denotes the closed linear space of $\{x_\alpha|\alpha\in\bigcup_{i=1}^n A_i\}$, then D_n is a closed proper subspace of X and therefore nowhere dense. Also $X=\bigcup_{n=1}^{\aleph_0}D_n$; but this contradicts the Baire category theorem [3, p. 121].

REMARK. The cardinal $\aleph_0$ is of type I, while the cardinality of the continuum is not.

Acknowledgments. The authors would like to thank Professor A. Wilansky for referring them to the work of Goffman [2] and suggesting its relation to Theorem 1. They would also like to thank Professors H. G. Burchard and J. B. Rosser for many helpful comments. The numerous conversations with Professor Burchard greatly facilitated the writing of this note.

Sponsored by the Mathematics Research Center, United States Army, Madison, Wisconsin, under Contract No. DA-31-124-ARO-D-462.

References

1. M. G. Arsove and R. E. Edwards, Generalized bases in topological linear spaces, Studia Math., 19 (1960) 95–113.

2. C. Goffman, On linear spaces which may be rendered complete metric spaces, Bull. Amer. Math. Soc., 49 (1943) 611–614.

3. ——— and G. Pedrick, First Course in Functional Analysis, Prentice-Hall, Englewood Cliffs, N. J., 1965.

4. A. H. Kruse, Badly incomplete normed linear spaces, Math. Z., 83 (1964) 314–320.

5. A. Markouchevitch, Sur les bases (au sens large) dans les espaces linéaires, Dokl. Akad. Nauk (N. S.), 41 (1943) 227–229.

6. A. Tarski, Quelques théorèmes sur les alephs, Fund. Math., 7 (1925) 1–14.

7. R. A. Tapia and J. W. Evans, Schauder dimension in Banach spaces, Mathematics Research Center Report #863, University of Wisconsin, 1968.

RESEARCH PROBLEMS

EDITED BY VICTOR KLEE

In this Department the Monthly presents easily stated research problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Material should be sent to Victor Klee, Department of Mathematics, University of Washington, Seattle, WA 98105.

TO WHAT LIMITS DO COMPLEX ITERATED RADICALS CONVERGE?

C. S. OGILVY, Hamilton College

Let $f(x)=\lim_{n\to\infty} f_n(x)$ be defined for all positive real x for which this limit exists, where $f_1(x)=\sqrt{x}$ and $f_n(x)=\sqrt{x+f_{n-1}(x)}$, $n>1$. That is, $f(x)$ is the limit of the *partial radicals* of the iterated radical

$$R(x) = \sqrt{x+\sqrt{x+\sqrt{x+\cdots}}}\,.$$

If $f(x)$ exists then it is also equal to $\lim_{n\to\infty} f_{n-1}(x)$, and we can write $y=f(x)=\sqrt{x+y}$, $x>0$. Solving this quadratic for y yields $y=\frac{1}{2}(1\pm\sqrt{1+4x})$, the parabola whose axis is $y=\frac{1}{2}$ with vertex at $(-\frac{1}{4},\frac{1}{2})$ and focus at $(0,\frac{1}{2})$. We can separate this into two functions, $y_1=\frac{1}{2}(1+\sqrt{1+4x})$ and $y_2=\frac{1}{2}(1-\sqrt{1+4x})$, the upper and lower "halves" of the parabola respectively.

It is well known that $R(x)$ converges to $y_1(x)$ for all $x>0$, [1]. But note that $R(0)=0=y_2(0)$.

The first problem is to determine the convergence of $R(x)$ for $-\frac{1}{4}\leqq x<0$, where the partial radicals $f_n(x)$ are all complex. We seek to prove (or disprove) the following conjecture, based on numerical calculation and plotting of the first few f_n:

1. *For* $-\frac{1}{4}\leqq x<0$, *$R(x)$ converges to the real value* $y_1(x)$, no matter whether the square root of a complex number is consistently taken as the one with the least positive amplitude or consistently taken as the other one.

Is there any iterated radical that converges to values given by $y_2(x)$, $x\neq 0$? A logical candidate is

$$Q(x) = -\sqrt{x-\sqrt{x-\sqrt{x-\cdots}}}\,,$$

because if Q has a limit y, then $y=-\sqrt{x+y}$, which leads to the same quadratic as before. It is clear that $Q(1)$ diverges.

Our next conjecture is

2. *$Q(x)$ converges to $y_2(x)$ for all $x>1$.*

It is easy to prove by induction that if for any $x<0$, $R(x)\to z$, then $Q(x)\to\bar{z}$, the complex conjugate of z. Thus if conjecture 1 is correct, then we know that the values of $Q(x)$ for $-\frac{1}{4}<x<0$ are given not by y_2 but, surprisingly, by y_1. In the range $0<x<1$, the partial radicals associated with Q are complex for all $n\geqq 2$, whereas the corresponding $f_n(x)$ are real, and the induction proof cannot be carried over. Nevertheless, it would still appear that

3. *For all $0<x<1$, $Q(x)\to y_1(x)$.* One may also ask:

4. *What iterated radical, if any, converges to $y_2(x)$ for $-\frac{1}{4}<x<0$ and $0<x<1$?* Finally, a broader objective would be:

5. *Develop procedures by which to investigate the convergence of more general iterated radicals, such as*

$$\sqrt{a_1+\sqrt{a_2+\sqrt{a_3+\cdots}}},$$

where the a_i are negative or complex. Herschfeld [1] remarks only that "Convergence questions appear to become very difficult in such cases."

Reference

1. Aaron Herschfeld, On infinite radicals, this MONTHLY, 42 (1935) 419.

ARE THERE AN INFINITY OF UNITARY PERFECT NUMBERS?

M. V. SUBBARAO, University of Alberta

A divisor d of n is said to be unitary if d and n/d are relatively prime. We shall denote the sum of the unitary divisors of n by $\sigma^*(n)$. Let n be called a unitary perfect number provided $\sigma^*(n)=2n$.

The inevitable question arises: are there an infinity of unitary perfect numbers? P. Erdös to whom the writer mentioned this problem in 1965, expressed the opinion that it might be a very difficult one, comparable to the problem of odd perfect numbers and he readily offered a prize of \$10 for the first complete solution, to which the writer offers a supplemental prize of an equal amount.

It is trivial to show that there are no odd unitary perfect numbers. Suppose n is an even unitary perfect number and is of the form $n=2^a m$, where m is odd and has r distinct prime divisors. In a paper in 1965, L. J. Warren and the writer showed, by elementary methods, the following theorem:

THEOREM A.

(i) *If $r=1$, then $n=60$.*
(ii) *If $a=1$, then $n=6$ or 90.*
(iii) *If $a=2$, then $n=60$.*
(iv) *If $r=2$, then $n=60$ or 90.*
(v) *a cannot be 3, 4, 5 or 7.*
(vi) *r cannot be 3 or 5.*
(vii) *If $a=6$, then $n=87{,}360$.*
(viii) *If $r=4$, then $n=87{,}360$.*

A year ago, the writer together with three undergraduate students, T. J. Cook, R. S. Newberry and J. M. Weber, (participants in the Undergraduate Research Participation Program under National Science Foundation Grant No. GY 4599 to the University of Missouri) obtained further results in this direction including the following:

THEOREM B. *Let* $n=2^a m$ *be unitary perfect. With the same notation as in Theorem* A, (i) *it is not possible for* $a=8, 9, 10$; (ii) *it is not possible for* $r=6$.

The proof involves extensive and exhausting calculations using a desk calculator. The details are too long to be shown here.

These theorems can be used to show, for example, that after 87,360, there can be no unitary perfect number with less than 20 digits. The writer was recently informed, however, that a graduate student at the University of Tennessee, Mr. Charles R. Wall, found by accident another unitary perfect number, namely

$$2^{18}\cdot 3\cdot 5^4\cdot 7\cdot 11\cdot 13\cdot 19\cdot 37\cdot 79\cdot 109\cdot 157\cdot 313,$$

a number with 24 digits!

Therefore it seems rather unsafe to make the conjecture that there are only a finite number of unitary perfect numbers, but the writer is still inclined to make it!

Some other results involving these numbers will be given elsewhere.

References

1. M. V. Subbarao and L. J. Warren, Unitary perfect numbers, Canadian Math. Bull., 9 (1966) 147–153.

2. M. V. Subbarao, T. J. Cook, R. S. Newberry, and J. M. Weber, On unitary perfect numbers, (to appear).

3. C. R. Wall. (Private communication from L. J. Warren.)

CLASSROOM NOTES

EDITED BY DAVID DRASIN

Manuscripts for this Department should be sent to David Drasin, Division of Mathematical Sciences, Purdue University, Lafayette, IN 47907.

ISOLATION OF ZEROS IN THE SECOND ORDER LINEAR DIFFERENTIAL EQUATION

M. R. CULLEN, Louisiana State University

Let $\psi(x)$ be a solution ($\psi \not\equiv 0$) to the second order linear differential equation

$$y'' + \alpha_1(x)y' + \alpha_2(x)y = \beta(x),$$

where $\alpha_1(x)$, $\alpha_2(x)$, and $\beta(x)$ are continuous on (a, b). When $\beta(x)\equiv 0$, it is well

known that the zeros of $\psi(x)$ are isolated, and this appears in many texts as an application of the fundamental uniqueness theorem [1, p. 105]. We wish to show that in case $\beta(x)$ has isolated zeros on (a, b), the same conclusion holds for $\psi(x)$. This result is already known. In the more general case of a nonhomogeneous nth order linear differential equation, the fact that $\psi(x)$ has isolated zeros may be deduced from a mean value theorem of Polya ([2], page 313, Theorem I) or from a related result of Widder ([3], page 394, Theorem III). The proof of this mean value theorem, although quite elegant, involves several manipulations with nth order Wronskians, and the theorem of Widder requires considerable preparation. The proof which will be presented (for the case $n=2$) is elementary, and should give the student in an intermediate differential equations course (such as in [1]) an opportunity to put to use some of the key theorems regarding second order linear differential equations (namely, the uniqueness theorem, Abel's formula, and the variation of parameters formula).

LEMMA 1. *Suppose $\beta(x)\equiv 0$ and that $\phi(x)$ is a nontrivial solution. Then the zeros of $\phi(x)$ are isolated.*

Proof. Suppose not. Then there exists an $x_0\in(a, b)$ and a sequence x_n in (a, b) such that $\phi(x_n)=0$ and $x_n\to x_0$. Hence, $\phi(x_0)=0$ and

$$\phi'(x_0) = \operatorname*{limit}_{n\to\infty} [\phi(x_n) - \phi(x_0)]/(x_n - x_0) = 0.$$

By the uniqueness theorem [1, p. 105], ϕ is identically zero, a contradiction.

LEMMA 2. *There exist linearly independent solutions $\phi_1(x)$ and $\phi_2(x)$ to the homogeneous differential equation such that $\phi_1(x)/\phi_2(x)$ is increasing between successive zeros of $\phi_2(x)$.*

Proof. Let $\phi_1(x)$ and $\phi_2(x)$ be two linearly independent solutions. For $\phi_2(x)\neq 0$,

$$[\phi_1(x)/\phi_2(x)]' = W(\phi_2, \phi_1; x)/[\phi_2(x)]^2.$$

By Abel's formula [1, p. 113], the Wronskian $W(\phi_2, \phi_1; x)$ is of constant sign on (a, b). Hence, replacing ϕ_1 by $-\phi_1$, if necessary, we may assume that the sign of the Wronskian is positive, and the result follows.

THEOREM. *Suppose $\psi(x)$ is a solution to the nonhomogeneous second order linear differential equation $y''+\alpha_1(x)y'+\alpha_2(x)y=\beta(x)$, where $\beta(x)$ has isolated zeros on (a, b). Then $\psi(x)$ has isolated zeros.*

Proof. Suppose not. Then, as in Lemma 1, there is an x_0 in (a, b) such that $\psi(x_0)=0$ and $\psi'(x_0)=0$. By the uniqueness theorem [1, p. 90, exercise 6], and the variation of parameters formula,

$$\psi(x) = \int_{x_0}^{x} \beta(t)[\phi_1(t)\phi_2(x) - \phi_2(t)\phi_1(x)]/W(t)dt,$$

where $W(t)=W(\phi_1, \phi_2; t)$ and ϕ_1 and ϕ_2 are as in Lemma 2. Since the zeros of $\beta(x)$ and $\phi_2(x)$ are isolated, we may choose $r>0$ such that $\beta(x)$ and $\phi_2(x)$ have no zeros on (x_0-r, x_0) and (x_0, x_0+r). First, consider $x\in(x_0, x_0+r)$. For $x_0<t<x$, we have $\phi_1(x)/\phi_2(x)>\phi_1(t)/\phi_2(t)$; and, multiplying by the positive number $\phi_2(x)\phi_2(t)$, we have

$$\phi_1(x)\phi_2(t) - \phi_2(x)\phi_1(t) > 0.$$

Hence the integrand is of one sign for $x_0<t<x$; and, therefore, $\psi(x)$ is nonzero. Similarly, for $x\in(x_0-r, x_0)$ and $x<t<x_0$, the integrand is of one sign, and hence $\psi(x)$ is nonzero. But we assumed x_0 was a limit point of zeros, a contradiction.

It is easy to see that the above theorem is false if $\beta(x)$ is not identically zero, yet has nonisolated zeros. Consider $y''=\beta(x)$, where $\beta(x)=x(20x^2-1)\sin(1/x) -8x^2\cos(1/x)$. $\beta(x)$ changes sign infinitely often about zero (consider $1/n\pi$), and $\psi(x)=x^5\sin(1/x)$ is a solution with $x=0$ as a nonisolated zero.

Considerably more can be said about the zeros of nontrivial solutions. For many other theorems of this type, the reader is referred to [2] and [3], where questions such as the number of zeros on an interval and interpolation by solutions are discussed.

References

1. E. A. Coddington, An Introduction to Ordinary Differential Equations, Prentice-Hall, Englewood Cliffs, New Jersey, 1961.

2. G. Polya, On the mean value theorem corresponding to a given linear homogeneous differential equation, Trans. Amer. Math. Soc., 24 (1922) 312–324.

3. D. V. Widder, A general mean value theorem, Trans. Amer. Math. Soc., 26 (1924) 385–394.

THE GENERALIZED JORDAN CANONICAL FORM

D. W. ROBINSON, Brigham Young University

One of the topics usually considered in a course in linear algebra is the study of various matrix representations of a linear transformation on a vector space over a field. If the field is algebraically closed, then the most useful and well-known representation is the Jordan canonical form. However, it is not as well known that, by means of a slight extension, this form may essentially be used over a larger class of fields. The purpose of this note is to suggest a way to bring this "generalized Jordan canonical form" into the classroom.

Let X be a vector space of finite dimension over a field K, and let τ be a linear transformation on X. First, suppose τ is cyclic with minimum polynomial π^r, where

$$\pi = \lambda^s - a_{s-1}\lambda^{s-1} - \cdots - a_1\lambda - a_0$$

is irreducible over K. Since τ is cyclic, there is a vector x in X such that

$$\{x, x\tau, \cdots, x\tau^{s-1}, x\pi(\tau), x\pi(\tau)\tau, \cdots, x\pi(\tau)\tau^{s-1}, \cdots, x\pi(\tau)^{r-1}\tau^{s-1}\}$$

is a basis of X. The matrix of τ relative to this basis is

$$H(\pi^r) = \begin{bmatrix} C & N & 0 \cdots 0 \\ 0 & C & N \cdots 0 \\ \cdot & \cdot & \cdot \cdots \cdot \\ 0 & 0 & 0 \cdots C \end{bmatrix}$$

where

$$C = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ a_0 & a_1 & a_2 & \cdots & a_{s-1} \end{bmatrix}, \qquad N = \begin{bmatrix} 0 & 0 \cdots 0 \\ \cdot & \cdot \cdots \cdot \\ 0 & 0 \cdots 0 \\ 1 & 0 \cdots 0 \end{bmatrix}.$$

$H(\pi^r)$ is called the hypercompanion matrix of π^r.

More generally, suppose τ has elementary divisors $\pi_i^{r_{ij}}$, $j=1, \cdots, m_i$, $i=1, \cdots, m$. It may be shown that X is decomposable into a direct sum of subspaces X_{ij}, where X_{ij} is invariant under τ and the restriction of τ to X_{ij} is cyclic with minimum polynomial $\pi_i^{r_{ij}}, j=1, \cdots, m_i, i=1, \cdots, m$. Thus, since a basis of X may be obtained by stringing together bases of the subspaces X_{ij}, it follows that τ may be represented by the direct sum of the hypercompanion matrices $H(\pi_i^{r_{ij}})$, $j=1, \cdots, m_i$, $i=1, \cdots, m$. (For further details see, for example, [2], p. 161–162.)

In case $s=1$, the hypercompanion matrix reduces to the usual Jordan matrix with $C=a_0$ and $N=1$. In this case the matrix representation enjoys a number of convenient computational advantages. For example, it is very easy to express the value of any power of the Jordan matrix. However, if $s>1$, then computational difficulties arise. Indeed, since C and N do not in general commute, it is awkward to even express the specific form of a power of $H(\pi^r)$.

On the other hand, if I is the s-by-s identity matrix, then

$$G(\pi^r) = \begin{bmatrix} C & I & 0 \cdots 0 \\ 0 & C & I \cdots 0 \\ \cdot & \cdot & \cdot \cdots \cdot \\ 0 & 0 & 0 \cdots C \end{bmatrix}$$

behaves in many ways as the usual Jordan matrix. For example, if t is a positive integer, then

$$(G(\pi^r))^t = \begin{bmatrix} C^t & \binom{t}{1} C^{t-1} & \binom{t}{2} C^{t-2} \cdots & \binom{t}{r-1} C^{t-(r-1)} \\ 0 & C^t & \binom{t}{1} C^{t-1} \cdots & \binom{t}{r-2} C^{t-(r-2)} \\ \cdot & \cdot & \cdot \cdots & \cdot \\ 0 & 0 & 0 \cdots & C^t \end{bmatrix}.$$

We call the rs-by-rs matrix $G(\pi^r)$ the generalized Jordan matrix corresponding to π^r. (The reader is referred to [3], where specific use is made of this particular matrix.)

The question arises, under what conditions are the matrices $H(\pi^r)$ and $G(\pi^r)$ similar? In other words, when is it legitimate to use $G(\pi^r)$ as the matrix representation of a cyclic linear transformation with minimum polynomial π^r? Since a matrix is cyclic if and only if the degree of its minimum polynomial is the order of the matrix, this question is answered by the following lemma, where π' denotes the derivative of π.

LEMMA. *If π is irreducible and $r>1$, then the generalized Jordan matrix $G(\pi^r)$ has minimum polynomial π^r if and only if $\pi' \neq 0$.*

Proof. Let $G=G(\pi^r)$ and note from the expression of G^t above that

$$\pi(G) = \begin{bmatrix} \pi(C) & \pi'(C) & * & \cdots & * \\ 0 & \pi(C) & \pi'(C) & \cdots & * \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & 0 & \cdots & \pi(C) \end{bmatrix}.$$

Hence, because $\pi(C)=0$,

$$(\pi(G))^{r-1} = \begin{bmatrix} 0 & 0 \cdots 0 & (\pi'(C))^{r-1} \\ 0 & 0 \cdots 0 & 0 \\ \cdot & \cdots & \cdot \\ 0 & 0 \cdots 0 & 0 \end{bmatrix}.$$

Now suppose $\pi'=0$. Then $(\pi(G))^{r-1}=0$ and π^r is not the minimum polynomial of G. Conversely, suppose $\pi'\neq 0$. Since π is irreducible, this means that π and π' are relatively prime, and it follows that $\pi'(C)$ is nonsingular. Thus, $(\pi(G))^{r-1}\neq 0$, $(\pi(G))^r=0$, and π^r is the minimum polynomial of G.

An irreducible polynomial with a nonzero derivative is said to be separable ([5], p. 65). Thus, as a consequence of this lemma and the observations above, we have the following (compare [1], p. 115):

THEOREM. *If τ is a linear transformation on a finite-dimensional vector space over a field, then τ may be represented by the direct sum of the generalized Jordan matrices corresponding to its elementary divisors if and only if every irreducible factor with multiplicity greater than one in the minimum polynomial of τ is separable.*

An arbitrary polynomial is said to be separable provided each of its irreducible factors is separable, and is said to be semisimple provided it is the product of distinct irreducible factors. If we call a linear transformation separable whenever its minimum polynomial is separable, and semisimple when-

ever its minimum polynomial is semisimple ([4], p. 678), then we have the following

COROLLARY. *Let τ be a linear transformation on a finite-dimensional vector space X. Then τ is separable if and only if there exist linear transformations σ and ρ on X such that σ is separable and semisimple, ρ is nilpotent, $\tau=\sigma+\rho$, and $\rho\sigma=\sigma\rho$.*

Proof. Suppose that each of the irreducible factors of the minimum polynomial of τ is separable. Then, by the theorem above, we may represent τ by the direct sum of generalized Jordan matrices. But, it is clear from the form of $G=G(\pi^r)$ that if $S=\operatorname{diag}(C, C, \cdots, C)$ and $R=G-S$, then S has minimum polynomial π, R is nilpotent, $G=S+R$, and $RS=SR$. Hence, by piecing these parts together, we may readily identify the required linear transformations σ and ρ.

Conversely, if such σ and ρ exist, then let $\mu=\pi_1 \cdots \pi_m$ be the minimum polynomial of σ, where each π_i is irreducible and separable. Since $\mu(\sigma)=0$, $\tau=\sigma+\rho$, and $\rho\sigma=\sigma\rho$, it follows that $\mu(\tau)=\mu(\sigma+\rho)=\mu(\sigma)+\omega(\rho, \sigma)\rho=\omega(\rho, \sigma)\rho$ for some $\omega(\rho, \sigma)$. Hence, since ρ is nilpotent, $\mu(\tau)$ is nilpotent and the minimum polynomial of τ divides a power of μ. In particular, since each π_i is separable, so also is each of the irreducible factors of the minimum polynomial of τ. That is, τ is separable.

The σ and ρ of this corollary are unique and are called, respectively, the semisimple and nilpotent parts of τ. (See, for example, [4], p. 679.) They may, of course, be constructed independent of the discussion above. This in turn provides an alternative approach to the discussion of the generalized Jordan form. Indeed, given this decomposition of τ, it is a simple matter to construct a basis of X, so that the matrix representation of τ is in the generalized Jordan form.

To illustrate this fact, we conclude this note by constructing such a basis in case τ is again cyclic with minimum polynomial π^r, where π is irreducible, separable, and of degree s. In this case, the semisimple part σ has minimum polynomial π, and the nilpotent part σ is nilpotent of index r. Moreover, since σ and ρ commute and $\tau=\sigma+\rho$, if x is such that X is spanned by $\{x\tau^i\}$, then X is also spanned by $\{x\sigma^i\rho^j\}$. Consequently,

$$\{x, x\sigma, \cdots, x\sigma^{s-1}, x\rho, x\sigma\rho, \cdots, x\sigma^{s-1}\rho, \cdots, x\sigma^{s-1}\rho^{r-1}\}$$

is a basis of X. Finally, it is easily checked that $G(\pi^r)$ is the matrix of τ relative to this basis.

References

1. A. I. Mal'cev, Foundations of Linear Algebra, Freeman, San Francisco, 1963.
2. S. Perlis, Theory of Matrices, Addison-Wesley, Cambridge, Mass., 1962.
3. D. W. Robinson, On matrix commutators of higher order, Canad. J. Math., 17 (1965) 527–532.
4. S. Warner, Modern Algebra, Vol. II, Prentice-Hall, Englewood Cliffs, N. J., 1965.
5. O. Zariski and P. Samuel, Commutative Algebra, Vol. I, Van Nostrand, Princeton, N. J., 1958.

AN ELEMENTARY PROOF THAT PRIMES ARE SCARCE

E. L. SPITZNAGEL, JR., Washington University

In teaching basic number theory to a freshman mathematics class, I became interested in finding a proof that the density of primes is $o(1)$ which would be intelligible to most of my students. I believe the proof sketched below meets this criterion. It starts from first principles and appears to be simpler than the one given by Hardy and Wright. As usual, we let $\pi(n)$ denote the number of primes which are $\leqq n$.

THEOREM. *Let C be any number >0. Then there is a natural number N such that $\pi(n)/n<C$ for all $n\geqq N$.*

Proof. For each natural number n, let i be the natural number such that $2^{2i-2}<n\leqq 2^{2i}$. Then

$$\pi(n)/n \leqq \pi(2^{2i})/n < \pi(2^{2i})/2^{2i-2} = 4\pi(2^{2i})/2^{2i}.$$

Let j be any natural number. Each prime p, $2^{j-1}<p\leqq 2^j$, is a divisor of the binomial coefficient

$$\binom{2^j}{2^{j-1}}.$$

From this fact we deduce the following series of inequalities:

$$2^{2^j} \geqq \binom{2^j}{2^{j-1}} \geqq \prod_{2^{j-1}<p\leqq 2^j} p \geqq (2^{j-1})^{\pi(2^j)-\pi(2^{j-1})}.$$

Now if we take the exponent of 2 on each end of this series, we find, for $j>1$,

$$(*) \qquad \pi(2^j) - \pi(2^{j-1}) \leqq 2^j/(j-1).$$

Upon replacing the denominator $j-1$ in (*) with the number 1, we find that the number of primes $p\leqq 2^i$ is less than or equal to $2^i+2^{i-1}+\cdots+2^2+1<2^{i+1}$. Also, if the denominator $j-1$ in (*) is replaced by the number i, it follows that the number of primes p, $2^i<p\leqq 2^{2i}$, is less than or equal to

$$(2^{2i}+2^{2i-1}+\cdots+2^{i+1})/i < 2^{2i+1}/i.$$

Therefore $\pi(2^{2i})<2^{i+1}+(2^{2i+1}/i)$, and thus

$$\pi(2^{2i})/2^{2i} < 1/2^{i-1} + 2/i.$$

Pick k to be the smallest natural number such that $1/2^{k-1}+2/k<C/4$. Then for all $i\geqq k$, $1/2^{i-1}+2/i<C/4$. Finally, set $N=2^{2k}$. If $n\geqq N$, then $i\geqq k$, and so

$$\pi(n)/n < 4\pi(2^{2i})/2^{2i} < 4C/4 = C,$$

as was to be shown.

Note that the proof makes no use of the concept of logarithm, with which, in the writer's experience, many high school graduates claim they are not familiar.

However, to someone who is familiar with logarithms, the proof actually yields $\pi(n)/n = O(1/\log n)$, though with a larger bounding constant than the one usually obtained in a proof of Tchebychev's Theorem.

I am indebted to Larry Dornhoff, who read an early version of this proof and suggested several improvements. In particular, his idea of using the two telescoping sums made the argument much more straightforward.

A NOTE CONCERNING THE LAW OF QUADRATIC RECIPROCITY

PETER HAGIS, JR., Temple University

If p and q are distinct odd primes we shall write pRq if p is a quadratic residue of q, and pNq if p is a quadratic nonresidue of q. For a given odd prime p we consider the problem of determining all the odd primes q such that pRq. Such problems are considered for special values of p in most texts on elementary number theory, and the results to be obtained here are implicit in the discussions of the applications of the Quadratic Reciprocity Law to be found in these books. I have not, however, seen an explicit statement of the following

THEOREM 1. *Let p and q be distinct odd primes and write $q = 2pk \pm r$, $0 < r < p$, if $p \equiv 1 \pmod 4$; and $q = 4pk \pm r$, $0 < r < 4p$, $r \equiv 1 \pmod 4$, if $p \equiv 3 \pmod 4$. Then pRq if and only if rRp.*

Proof. If (r/p) is the Legendre symbol we recall that $(r/p) = (s/p)$ if $r \equiv s \pmod p$, and $(-r/p) = \pm(r/p)$ according as $p \equiv \pm 1 \pmod 4$. Therefore, if $p \equiv 1 \pmod 4$ we have from the Law of Quadratic Reciprocity $(p/q) = (\pm r/p) = (r/p)$ from which the conclusion of the theorem is immediate. If $p \equiv 3 \pmod 4$ then, since $r \equiv 1 \pmod 4$, $(p/(4pk+r)) = (r/p)$. Also, since $-r \equiv 3 \pmod 4$ we have $(p/(4pk-r)) = -(-r/p) = (r/p)$. This completes the proof.

We give two examples. If $p = 13$ then the least positive odd quadratic residues of 13 are 1, 3, 9. Therefore, $13Rq$ if q is of the form $26k \pm r$ where $r = 1, 3, 9$, and $13Nq$ otherwise. If $p = 7$ then the positive odd quadratic residues of 7 less than 28 and congruent to 1 modulo 4 are 1, 9, 25. Therefore, $7Rq$ if and only if $q = 28k \pm r$, $r = 1, 9, 25$.

A similar result holds for $-p$. For we have

THEOREM 2. *Let p and q be distinct odd primes and write $q = 2pk + r$, $0 < r < 2p$, if $p \equiv 3 \pmod 4$; and $q = 4pk + r$, $0 < r < 4p$, if $p \equiv 1 \pmod 4$. If $p \equiv 3 \pmod 4$ then $-pRq$ if and only if rRp. If $p \equiv 1 \pmod 4$ then $-pRq$ if and only if $r \equiv 1 \pmod 4$ and rRp, or $r \equiv 3 \pmod 4$ and rNp.*

Proof. We again employ the Law of Quadratic Reciprocity. Assume first that $p \equiv 3 \pmod 4$. If $q \equiv 1 \pmod 4$ then $(-p/q) = (p/q) = (r/p)$, while if $q \equiv 3 \pmod 4$ then $(-p/q) = -(p/q) = (r/p)$. If $p \equiv r \equiv 1 \pmod 4$ then $(-p/q) = (p/q) = (r/p)$. If $p \equiv 1 \pmod 4$ and $r \equiv 3 \pmod 4$ then $(-p/q) = -(p/q) = -(r/p)$.

For example, if $p=7$ then the positive odd quadratic residues of 7 less than 14 are 1, 9, 11. Therefore, $-7Rq$ if and only if $q=14k+r$ where $r=1, 9, 11$. If $p=5$ then the odd quadratic residues of 5 less than 20 and congruent to 1 modulo 4 are 1, 9, while the odd nonresidues less than 20 and congruent to 3 modulo 4 are 3, 7. Therefore, $-5Rq$ if and only if $q=20k+r$, $r=1, 3, 7, 9$.

MATHEMATICAL EDUCATION

EDITED BY J. G. HARVEY AND M. W. POWNALL

Material for this Department should be sent to either of the editors: J. G. Harvey, Department of Mathematics, University of Wisconsin, Madison, WI 53706; M. W. Pownall, Department of Mathematics, Colgate University, Hamilton, NY 13346.

PANEL DISCUSSION ON ASSISTANCE TO DEVELOPING COLLEGES

LLAYRON L. CLARKSON, Texas Southern University

The purposes of this presentation are to give an overview of some of the major problems in the undergraduate mathematics programs in developing institutions, with particular emphasis on the so-called Black institutions, and to make some recommendations to the Association.

Brief narrative. It is obvious that the problems and programs in the predominantly Black institutions are not fully understood by the major organizations. For example, the problem of what to do about the mathematics education of the nonscience major continues to cause great concern in the Black institutions. To my knowledge, none of these institutions has decided not to teach mathematics to these students. Yet we recognize that, to a large extent, the remedial and compensatory programs we conduct either duplicate or replace the high schools' mathematics instructional programs. But what are we to do? Many students do not have the basic mathematical skills necessary to complete successfully nonmathematics courses where mathematics is used. Many want to do graduate work in fields requiring some use of mathematics. Many accept teaching positions and many others take examinations requiring some elementary mathematics.

Many programs seeking better ways to communicate mathematics to this segment of the student population are now being conducted. The Association should call a series of conferences for the purposes of cataloguing, analyzing, and evaluating the present remedial and compensatory type programs. Such a conference might want to recommend to the Association that it should form an *ad hoc* curriculum committee for the specific purpose of considering curriculum

materials for nonscience majors in developing institutions; indeed all institutions that accept students who are academically deficient.

Of course, there are problems in designing efficient programs for majors in the various sciences and mathematics. Although many of these students are relatively competent in the so-called basic skills, they have developed many antiquated patterns and they have poor ideas about the true nature of mathematics and about the value and use of mathematics in today's society. As a consequence, the well-prepared and well-motivated freshman has to be counseled carefully and frequently. Otherwise, his first real mathematical experience at the college level could cause him extreme frustration, and it could cause the science community to lose forever a potential professional mathematician.

For further revelations into these problems I recommend that the Association form a task force with strong Black representation to catalogue the needs in the mathematics education programs in Black institutions for the general enlightenment of the membership at large and to be used specifically in the formulation of policy, the designation of priorities in the Association.

If the major organizations are really interested in assisting the Black colleges, they can provide a real service by demanding that the Black schools be properly represented when legislation and guidelines for support programs for higher education are being drafted. Upon implementing such demands, measures should be taken to assure that well-informed spokesmen are participating, so that a reasonable portion of the funds will be allocated for training the disadvantaged, including Blacks, and for meeting other concerns of the developing institutions. Let me cite some examples of what I call the misallocation of funds. Many institutions that were legally segregated just a few years ago or that are physically remote from major Black communities are receiving large grants to study Blacks. In my opinion, this practice is not a product of sound judgment; rather it is a product of political manipulations. Many of these institutions have no feeling for the importance of such undertakings, nor the expertise and personnel for such investigations. Indeed, many of these grantee institutions must employ Blacks to implement their programs. It seems to me that if a real impact is to be made on many of the Black community problems plaguing the nation, the Black institutions that are close to these problems must be awarded a share of the leadership and the support necessary to do the research, to design new programs, and to supervise the implementation of these programs. It seems that the Black institutions are in a much better position to provide a good answer to, "Relevant to What?"

These are general remarks but there are significant implications to the mathematical community. I think that it is time for the mathematical community to broaden its scope and give rewards to its members who design humanistic applications of the subject matter. For example, current statistics indicate that only one Black out of 200 goes to college. At Texas Southern, only about 3 out of every 10 who enter finish with some kind of bachelor's degree. Can the mathe-

matical community stimulate more youngsters into higher education? Can we provide positive, valuable experiences so as to make better citizens of those who drop out? And how can we improve the quality of those who finish? These are all questions that must be answered, and they can be answered by the Black institutions provided they gain adequate support to do the necessary work.

The problem of training the next generation of Black mathematics professors must be considered. If the present trend continues, the call for Blacks for practically all universities and colleges will expand. I am guessing that the real talent among Blacks has only been partially tapped. I know that the necessary recruiting, and at least the initial training of many Blacks, can best be done in an atmosphere requiring the least amount of personal adjustment for these students. The Black institutions should be given the leadership in providing the undergraduate training for the next generation of Black teachers. There are significant implications here for the involvement of the mathematical community.

The final point I wish to make before I summarize my recommendations is that one day, very soon I hope, many veterans are going to return from Vietnam and other foreign lands. A significant number of these will be Black. Indications are that they will be seeking better lives for themselves and their families. I understand that the federal government has already considered making some special eligibility considerations for these veterans. Many of these veterans will need some precollege-level work in an adult setting. I am convinced that the Black schools are better equipped to serve the veterans. I am also convinced that the major universities will be awarded the significant grants to serve the country in this special way, unless major organizations such as the Association express realistic views on the proper utilization of all the national resources.

In addition to the problems briefly considered above, many predominantly Black institutions are experiencing significant changes that will affect their instructional programs. For example, many well-trained Black professors are leaving Black institutions to join the staffs of well-established institutions; many bright young students who would probably attend Black schools are attending large white institutions; a large number of well-trained nonblacks are actually seeking teaching positions in Black institutions; and many Black institutions are expressing commitments to deal directly with many community problems. There are numerous other examples. One can only speculate on the implications of these changes.

Recommendations. The following recommendations to the Association have been discussed with some of my colleagues:

1. We recommend that the Association call a series of two-day regional conferences for the specific purpose of cataloguing the problems and programs of Black institutions, including special programs such as those of the Institute for Services to Education and The Southern Education Foundation.

2. We recommend that the Association appoint an *ad hoc* curriculum committee to recommend to developing institutions meaningful alternatives to the current Committee on the Undergraduate Program in Mathematics undergraduate curriculum recommendations.
3. We recommend that the Association adopt as policy a motion stating that adequate representation from Black institutions, and indeed from developing institutions, be included in initial planning sessions and subsequent sessions when important issues concerning the obligations of the Association are to be considered.
4. We recommend that the Association use its power to influence congressmen, funding agencies, etc., to include representation from Black institutions in the initial sessions where legislation and guidelines for various programs concerning mathematics education are being considered.
5. We recommend that the Association survey present efforts for training the next generation of Black mathematics professors and then appoint an *ad hoc* committee with the specific charge of recommending to government, industry, funding agencies, etc., new ways to get more young Blacks involved.
6. We recommend that the Association research the needs in developing institutions concerning a variety of instruction-related problems, such as adequate graduate faculty, graduate fellowships, staff and student recruitment, travel expenses for professional activities, upgrading libraries, student assistants, visiting professors and lecturers, development of applied mathematics offerings, computers in education, precollege centers, curriculum development and other professional activities, research in learning and teaching mathematics, proper physical facilities, research instructorships, etc.
7. We recommend that the Association do what it can to stop the negative publicity the developing institutions, particularly the Black institutions, are getting; instead announce that these institutions are valuable assets to the national education network, that they do serve the communities, that they do conduct some basic research, that they do serve to provide educational opportunities to a significant segment of the population who would otherwise be denied such an opportunity and that they do turn out educated citizens. What more can one ask?

This article is based on a presentation to the Mathematical Association of America, Eugene, Oregon, August 1969.

PROBLEMS AND SOLUTIONS

EDITED BY EMORY P. STARKE

All problems (both elementary and advanced) proposed for inclusion in this Department should be sent to E. P. Starke, 1000 Kensington Ave., Plainfield NJ 07060. Proposers of problems are urged to enclose any solutions or information that will assist the editors. Ordinarily, problems in well-known textbooks and results in generally accessible sources are not appropriate for this Department. No solution (except those accompanying proposals) should be sent to Professor Starke.

ELEMENTARY PROBLEMS

Solutions of Elementary Problems should be sent to Problems Group, Mathematics Department, University of Maine, Orono, ME 04473. To facilitate their consideration, solutions of Elementary Problems in this issue should be typed (with double spacing) and should be mailed before July 31, 1970. Contributors (in the United States) who desire acknowledgment of receipt of their solutions are asked to enclose self-addressed stamped postcards.
An asterisk (*) means that no solution was submitted.

E 2208. [1970, 78]. **Correction.** *Proposed by D. Rameswar Rao, Secunderabad, India*

Let $n \geqq 5$ and $2 \leqq b \leqq n$. Prove

$$\left[\frac{(n-1)!}{b}\right] \equiv 0 \pmod{b-1}.$$

E 2228. *Proposed by C. C. Lindner, Emory University*

Call a latin square of order n an X_n-latin square provided that each of the n symbols on which it is based occurs on each of the two diagonals. Show that if $n = 2^k$, $k \geqq 2$, there exist $n-2$ mutually orthogonal X_n-latin squares.

E 2229. *Proposed by J. V. Michalowicz, The Catholic University of America*

Any two positive integers a, b have a g.c.d.(a, b) and a l.c.m.$[a, b]$, where $[a, b] = ab/(a, b)$. Can this be generalized to express the l.c.m. of any finite numbers of elements in terms of g.c.d.'s only?

E 2230. *Proposed by B. E. Rodden, Defence Research Board, Toronto, Canada*

(A) Urns I and II each contain exactly n balls numbered consecutively from 1 to n. One ball is drawn from urn I. Balls are then drawn one at a time from urn II until the same-numbered ball is found. No balls are replaced in the urns. The process is repeated until urn II is empty. Find p_{rn}, the probability of precisely r matches.

(B) For a more complex problem, m balls are drawn from urn I. Again balls are drawn one at a time from urn II and compared with the balls from urn I. When a match is made, one new ball is drawn from urn I to replace the matched ball, and the process is continued. When urn I is empty, the process continues with the remaining unmatched balls until urn II is empty. Find p_{rn}, the probability of precisely r matches in the case $m=2$.

E 2231.* *Proposed by M. S. Klamkin, Ford Scientific Laboratory*

It is a known result that if the centroid of the vertices and the centroid of the area (both uniformly weighted) of a quadrilateral coincide, then the figure is a parallelogram. If the centroids of the vertices, of the edges, and of the area (all uniformly weighted) of a pentagon all coincide, must the figure be a regular pentagon?

E 2232. *Proposed by H. D. Ruderman, Hunter College High School, New York City*

Let U_n be the smallest number of different unit fractions totalling 1 where the largest unit fraction is $\leq 1/n$. For example, $U_1=1$, $U_2=3$, and $U_3=5$ because

$$1=\frac{1}{3}+\frac{1}{4}+\frac{1}{5}+\frac{1}{6}+\frac{1}{20}\cdot$$

It is well known that for every n there is a finite number of terms giving the sum 1. Find an upper bound on U_n.

E 2233. *Proposed by E. J. Cockayne, University of Victoria, Canada*

Let A, B, C be three distinct points on the surface of a sphere, not all on the same great circle. The closed curve Γ formed by the minor great circle arcs AB, BC, CA divides the surface into two unequal areas. Suppose Z is the set of points which comprise the smaller area including the boundary Γ. Prove that any point P of the surface minimizing the sum of the minor great circle arcs $PA+PB+PC$ is a point of Z.

E 2234. *Proposed by P. M. Gibson, University of Alabama at Huntsville.*

Prove or disprove: If $A=(a_{ij})$ is a nonsingular $n\times n$ complex matrix, if also $A^{-1}=(b_{ij})$, and each a_{ij} and b_{ij} is nonzero, then the matrices of reciprocals (a_{ij}^{-1}) and (b_{ij}^{-1}) are singular or nonsingular together. H. Flanders [this MONTHLY, (1966) 270–272] proved this for $n=3$.

SOLUTIONS OF ELEMENTARY PROBLEMS

A Number Theory Problem

E 2142 [1969, 82; 1070]. *Proposed by Erwin Just, Bronx (N. Y.) Community College*

Let k, b and r be fixed integers. Call an integer n *special* if each member of $\{kb^n+i\}$, $i=1, 2, \cdots, r$, is composite. Prove that the number of special integers is infinite.

II. *Notes by the proposer.* (A) In the original solution, if $kb+i$ does not have an odd prime divisor, then Fermat's theorem, which is used in the proof, may not be applied. In this case, however, $kb+i$ is a power of 2 and kb^n+i is also even, hence composite.

(B) The result may be somewhat generalized. For arbitrary positive integers r and s, let $\{k_j\}$, $\{b_j\}$, $\{d_i\}$, $(i=1, 2, \cdots, r)$, $(j=1, 2, \cdots, s)$, be sets of fixed nonzero integers with $b_i\neq 1$. Call an integer n *special* if each member of $\{d_i+\sum_{j=1}^{s} k_jb_j^n\}$, $(i=1, 2, \cdots, r)$, is composite. The number of special integers is infinite.

The proof employs the techniques of the original solution. By choosing p_i to be a prime divisor of $d_i+\sum_{j=1}^{s} k_jb_j$ we see that, for any positive integer m, since

$$d_i + \sum_{j=1}^{s} k_jb_j^{1+m(p_i-1)} = \left(d_i + \sum_{j=1}^{s} k_jb_j\right) + \sum_{j=1}^{s} k_jb_j(b_j^{m(p_i-1)} - 1),$$

it must be the case that $d_i+\sum_{j=1}^{s} k_jb_j^{1+m(p_i-1)}$ is divisible by p_i. The required conclusion is obtained by choosing $n=1+m\prod_{i=1}^{r}(p_i-1)$.

III. *Comment by M. J. DeLeon, Florida Atlantic University.* The first sentence in parentheses on the second line of the solution needs modification. Replace "This is possible for any i unless $b|i$." by "This is possible for any i unless $(b, i)\neq 1$." For a counterexample to the original statement consider $k=1$, $b=45$, $i=1830$.

E 2143 [1969, 82, 413, 825; 1970, 79]. **Addendum.** Solutions to this problem were also contributed by C. S. Karuppan Chatty (India), G. R. Padmanaban (India), P. Nagasundaram (India), and H. R. van der Vaart.

A Basic Digital Inequality

E 2179 [1969, 690]. *Proposed by M. G. Beumer, Technological University, Delft, Netherlands*

If $n\geqq 0$, $r>1$, $0<a\leqq r$ (n, r, a integers), show that n has, in the scale of r, exactly

$$\sum_{k=1}^{\infty}\{[nr^{-k}+ar^{-1}]-[nr^{-k}]\}$$

digits which are $\geqq r-a$.

Solution by Leonard Carlitz, Duke University. Put

$$n = n_1 + n_2r + n_3r^2 + \cdots \qquad (0 \leqq n_k < r).$$

Then

$$n_1 = n - \left[\frac{n}{r}\right]r, \qquad n_2 = \left[\frac{n}{r}\right] - \left[\frac{n}{r^2}\right]r,$$

and generally $n_k=[n/r^{k-1}]-[n/r^k]r$ $(k=1, 2, 3, \cdots,)$. Clearly, for fixed k, $n_k\geqq r-a$ if and only if $[n/r^{k-1}]-[n/r^k]r\geqq r-a$, that is,

$$\left[\frac{1}{r}\left(\left[\frac{n}{r^{k-1}}\right]+a\right)\right] \geqq \left[\frac{n}{r^k}\right]+1.$$

But, if $n=mr^{k-1}+s$, $0 \leqq s < r^{k-1}$, then

$$\left[\frac{1}{r}\left(\left[\frac{n}{r^{k-1}}\right]+a\right)\right] = \left[\frac{m}{r}+\frac{a}{r}\right] = \left[\frac{m}{r}+\frac{s}{r^k}+\frac{a}{r}\right] = \left[\frac{n}{r^k}+\frac{a}{r}\right].$$

Thus $n_k \geqq r-a$ if and only if

$$\left[\frac{n}{r^k}+\frac{a}{r}\right] > \left[\frac{n}{r^k}\right].$$

On the other hand

$$0 \leqq \left[\frac{n}{r^k}+\frac{a}{r}\right] - \left[\frac{n}{r^k}\right] \leqq 1.$$

Hence

$$\left[\frac{n}{r^k}+\frac{a}{r}\right] - \left[\frac{n}{r^k}\right] = \begin{cases} 1 & (n_k \geqq r-a) \\ 0 & (n_k < r-a), \end{cases}$$

and so

$$\sum_{k=1}^{\infty}\left\{\left[\frac{n}{r^k}+\frac{a}{r}\right] - \left[\frac{n}{r^k}\right]\right\}$$

is equal to the number of digits n_k such that $n_k \geqq r-a$.

REMARK. The above argument evidently yields the following more general result. Let $0 < a_k \leqq r$ $(k=1, 2, 3, \cdots)$, and

$$n = n_1 + n_2 r + n_3 r^2 + \cdots \qquad (0 \leqq n_k < r).$$

Then the number of digits n_k such that $n_k \geqq r-a_k$ $(k=1, 2, \cdots)$ is equal to

$$\sum_{k=1}^{\infty}\left\{\left[\frac{n}{r^k}+\frac{a_k}{r}\right] - \left[\frac{n}{r^k}\right]\right\}.$$

Also solved by Anders Bager (Denmark), W. E. Donovan, Neal Felsinger, M. G. Greening, (Australia), D. C. B. Marsh, Simeon Reich (Israel), E. F. Schmeichel, Charles Wexler, and the proposer.

A Trigonometric Identity

E 2180 [1969, 691]. *Proposed by Norman Schaumberger, Bronx Community College, New York*

Show that $\sum_{j=0}^{n-1}(-1)^j \cos^n(\pi j/n) = n/2^{n-1}$.

Solution by Henry Ricardo, Yeshiva University and Manhattan College. Setting $\omega = \exp(\pi i/n)$, we have

$$\begin{aligned}\sum_{j=0}^{n-1}(-1)^j\cos^n(\pi j/n) &= \sum_{j=0}^{n-1}\omega^{nj}[(\omega^j+\omega^{-j})/2]^n\\ &= (1/2^n)\sum_{j=0}^{n-1}\omega^{nj}\sum_{k=0}^{n}\binom{n}{k}\omega^{j(n-2k)}\\ &= (1/2^n)\sum_{k=0}^{n}\binom{n}{k}\sum_{j=0}^{n-1}\omega^{j(2n-2k)}.\end{aligned}$$

Now $\sum_{j=0}^{n-1}\omega^{j(2n-2k)}=0$ unless $k=0$ or $k=n$, in which cases this sum is equal to n. Thus the original sum becomes

$$\frac{1}{2^n}\left[n\binom{n}{0}+n\binom{n}{n}\right]=n/2^{n-1}.$$

Also solved by Günter Bach (Germany), Anders Bager (Denmark), M. T. Bird, L. Carlitz, John Cerullo, M. S. Demos, G. C. Dodds, W. E. Donovan, W. O. Egerland, N. Ersec, O. H Fraser, R. Garfield, I. J. Good, M. G. Greening (Australia), J. E. Hafstrom, Steve Hartman, Philip Haverstick, Dennis Henkel, Robert Heller, Sidney Heller, A. C. Hindmarsh, Stephen Hoffman, John Horvâth, W. C. Huffman, C. M. Jensen, I. N. Katz, P. G. Kirmser, Lew Kowarski, J. R. Kuttler, O. P. Lossers (Netherlands), Beatriz Margolis (Argentina), D. C. B. Marsh, D. E. Nixon, Tj. Plomp (Netherlands), E. A. Power (Australia), Simeon Reich (Israel), Steve Rohde, Perry Scheinok, E. F. Schmeichel, J. S. Shipman, C. V. L. Smith, Dragutin Surtan (Yugoslavia), M. W. Wilson, M. R. Wise, Chung-liu Wong, and the proposer.

Several solvers gave an analogous identity involving the sine function. Good proved the following generalization

$$\sum_{j=0}^{n-1}(-1)^j\left(\cos\frac{\pi j}{n}\right)^n\cos\frac{2\pi js}{n}=\begin{cases}\dfrac{n}{2^n}\dbinom{n}{s} & \text{if } 1\leqq s\leqq n-1\\ \dfrac{n}{2^{n-1}} & \text{if } s=0,\end{cases}$$

making use of the fast Fourier transform, the result being equivalent to a note he contributed to *Nature*, vol. 222, p. 1302.

Equilateral Triangles on a Given Triangle

E 2181 [1969, 691]. *Proposed by Jack Garfunkel, Forest Hills High School, New York*

Given any triangle ABC and a given segment BP on side BC, determine (by geometric construction) segments CQ, AT on sides CA and AB respectively, so that equilateral triangles erected outwardly on these three segments have vertices that are the vertices of an equilateral triangle.

Solution by S. L. Greitzer, Rutgers—The State University. Let X be the third vertex of equilateral triangle BPX. Construct equilateral triangle XCC' and angle $BAZ=60°$. Let $C'Z$, parallel to CA, meet line AZ, thus determining Z, and draw line XZ. Then $AZ=AT$, and XZ is one side of the desired equilateral triangle.

Proof. Rotate triangle $XC'Z$ clockwise 60° about X. Then XC' will fall on

XC and $C'Z$ will take the position CY. So triangle ZXY is equilateral, and $CY=CQ$.

Also solved by Bernhard Andersen (Denmark), Lucio Artiaga, C. W. Eliason, Jr., Michael Goldberg, Norman Miller, A. T. Olson, W. W. Parsons, Tj. Plomp (Netherlands), Edna M. Pratt, Simeon Reich (Israel), E. F. Schmeichel, A. W. Walker, Mark Yu, and the proposer.

K-Sequences with Small Divisors

E 2182 [1969, 691]. *Proposed by N. P. Salz, Bangkok, Thailand*

Let a K-sequence be a block of K consecutive odd integers, each of which is divisible by at least one of the n odd primes, $p_1=3$, $p_2=5, \cdots, p_{n-1}, p_n$. Prove or disprove: If p_{n-2} divides at most one term of a K-sequence, then $K\leqq p_{n-1}-1$.

I. *Solution by Don Coppersmith, Massachusetts Institute of Technology.* The statement in general is untrue. The 19 consecutive odd integers starting with 82370091 are divisible by 3, 13, 5, 3, 7, 11, 3, 5, 23, 3, 19, 7, 3, 17, 13, 3, 11, 5, 3 respectively. Only one is divisible by 17. Here $n=8$, $K=19=p_{n-1}$.

Also the 32 consecutive odd integers beginning with 5546972216583 are divisible by 3, 5, 23, 3, 13, 17, 3, 7, 11, 3, 19, 5, 3, 29, 7, 3, 5, 13, 3, 11, 31, 3, 17, 37, 3, 23, 5, 3, 7, 19, 3, 5, respectively. Here only one is divisible by 29; $n=11$, $K=32=p_{n-1}+1$.

II. *Solution by Alfred Brauer, Wake Forest University.* In order to prove the existence of infinitely many primes in the arithmetic progression $kx+l$ with $(k, l)=1$, Legendre [1] tried to prove the following statement: Let p_n denote the nth prime. There exist at most $p_{n-1}-1$ consecutive odd integers such that each of them is divisible by at least one of the primes $p_2, p_3, \cdots, p_n$. Dirichlet [2] was the first to recognize that Legendre's proof was incorrect. He tried in vain to find out if the statement itself was correct or not. Hence, he did not use it for his proof for the theorem of the primes in arithmetic progression. In 1858, the Academy of Paris proposed for the *Grand Prix*: Prove or disprove the statement of Legendre. Five papers [3] were sent in, but none of them solved the problem completely. The prize was given to A. Dupré [4] who supposedly proved that the proposition is correct for $p_n=19$ and for $p_n=29$, 31, and 37, but wrong for $43\leqq p_n\leqq 113$. In 1930 H. Zeitz and I [5] finally proved that the proposition is incorrect for all primes $p_n\geqq 43$. Moreover, we proved the following theorem there:

For every $\epsilon>0$ there exists a number $n_0=n_0(\epsilon)$ such that for all $n>n_0$ there are sets of $(2-\epsilon)p_n$ consecutive odd integers of which no one is relatively prime to $P_n=p_2p_3\cdots p_n$.

We take $\epsilon=\frac{1}{2}$. Then there exists a set S_n of $[\frac{3}{2}p_n]$ such consecutive odd integers for $n>n_0$. Therefore the statement of the problem is incorrect for all those sufficiently large p_n for which S_n contains only one multiple of p_{n-2}.

Assume now that S_n contains at least two odd multiples of p_{n-2}. Let kp_{n-2} and $(k+2)p_{n-2}$ be the two smallest of them. Denote the smallest element of S_n by a and the greatest element of S_n which is smaller than $(k+4)p_{n-2}$ by b.

First we assume that at least $[\frac{1}{4}p_{n-2}]$ elements of S_n are less than kp_{n-2}. Then there are at least $[\frac{5}{4}p_{n-2}]$ elements of S_n less than $(k+2)p_{n-2}$, and only one of these, namely kp_{n-2}, is divisible by p_{n-2}. We denote this subset of S_n by T. Assume now that less than $[\frac{1}{4}p_{n-2}]$ elements of S_n are less than kp_{n-2}. Then more than $[\frac{1}{4}p_{n-2}]$ elements of S_n are greater than $(k+2)p_{n-2}$ since otherwise the number of elements of S_n would be less than

$$[\tfrac{1}{4}p_{n-2}] + p_{n-2} + [\tfrac{1}{4}p_{n-2}] \leqq [\tfrac{3}{2}p_{n-2}] < [\tfrac{3}{2}p_n].$$

This gives a contradiction. Therefore there are at least

$$p_{n-2} + [\tfrac{1}{4}p_{n-2}] = [\tfrac{5}{4}p_{n-2}]$$

elements of S_n in the closed interval $\{kp_{n-2}+2, b\}$ of S_n and only one of them is divisible by p_{n-2}, namely $(k+2)p_{n-2}$. We denote this subset of S_n by U. Then either T or U has at least $[\frac{5}{4}p_{n-2}]$ elements of which only one is divisible by p_{n-2}.

In order to prove that the statement of the problem is incorrect for all sufficiently large n we only have to show that

$$[\tfrac{5}{4}p_{n-2}] \geqq p_{n-1}.$$

It follows from the theorem of Tschebycheff for $x = p_{n-2}$ and $\delta = \frac{1}{4}$ that there exists at least one prime between p_{n-2} and $[\frac{5}{4}p_{n-2}]$ and p_{n-1} is the smallest of them.

References

1. A. M. Legendre, *Théorie des nombres*, (Paris, 1830), v. II, 71–79.

2. G. L. Dirichlet, *Beweis des Satzes, dass jede arithmetische Progression* · · · , Abhandl. Preuss. Akad. Wiss., (1837) 45–81 = Werke I, 313–342.

3. Rapport sur le concours pour le grand prix de sciences mathématiques, Comptes Rendus 48 (1859), 487–488.

4. A. Dupré, *Examen d'une proposition de Legendre relative à la théorie des nombres*, Paris, 1859, Mallet-Bachelier.

5. A. Brauer and H. Zeitz, *Über eine zahlentheoretische Behauptung von Legendre*, S.-B. Berlin. Math. Ges., 29 (1930), 116–125.

Also solved by D. E. Kuhn and by W. D. Maurer.

Editorial Note. It will be seen that Coppersmith's second set above provides a counterexample to Dupré's statement for $p_n = 37$.

Admissible Lines

E 2183 [1969, 691]. *Proposed by Steven Silverman, University of British Columbia*

We say that two points in the plane determine an admissible line if the line they determine is vertical, horizontal, has slope 1, or has slope -1. What is the maximum number of admissible lines n points can determine?

Solution by Sidney Heller and Larry Padwa, Brookhaven National Laboratory. Consider the n points to be a graph of n vertices. The maximum degree of each vertex is 4, one for each direction. Hence, by a fundamental theorem of graph theory, the sum of the degrees, $4n$, equals twice the number of edges in the graph. If no three vertices are collinear, then $2n$ is the upper limit for the number of admissible lines. Call a graph which attains this maximum *full.*

We know also that there are at most $\frac{1}{2}n$ edges in each direction which gives $2n-2$ admissible lines as the upper limit for n odd.

We can demonstrate that the full graph for any even n greater than or equal to 12 exists by exhibiting the full graph of order 8 and the full graphs of orders 12, 14, 18. For odd n greater than 12, the maximum of $2n-2$ admissible lines can be given by adding an isolated vertex to the full graph of even order $n-1$. Further, the upper limit of $2n-2$ is achieved for all odd n greater than 4.

Let $f(n)$ be the maximum number of admissible lines for n points. $f(2)=1$, $f(3)=3$, and $f(4)=6$, since the complete graph is the best obtainable, but of course it is not full. It can be proved that the full graphs of order 6 and 10 do not exist, but that $f(6)=11$ and $f(10)=19$. (The determination, particularly for $n=10$, is tedious and lengthy.)

To summarize: except for $n=2, 3, 4, 6, 10$, we have $f(n)=4\left[\frac{1}{2}n\right]$, and the exceptional values have been mentioned above.

Also solved by D. M. Bloom. Partial solutions by Don Coppersmith, R. B. Eggleton (Australia), R. W. Kehr & E. F. Schmeichel, O. P. Lossers (Netherlands), L. F. Meyers, E. C. Milner, C. C. Oursler, and Simeon Reich (Israel).

Note. The case $n=14$ seemed most troublesome. The full figure can be formed by taking two octagons with a common vertex removed.

ADVANCED PROBLEMS

All solutions of Advanced Problems should be sent to J. Barlaz, Rutgers—The State University, New Brunswick, NJ 08903. Solutions of Advanced Problems in this issue should be typed (with double spacing) on separate, signed sheets and should be mailed before July 31, 1970. Contributors (in the United States) who desire acknowledgment of receipt of their solutions are asked to enclose self-addressed stamped postcards.

5727. *Proposed by P. R. Halmos, University of Hawaii*

A matrix

$$M = \begin{vmatrix} A & B \\ C & D \end{vmatrix}$$

whose entries are matrices has four formal determinants: $AD-BC$, $AD-CB$, $DA-BC$, $DA-CB$. (In this sense a matrix of size n, instead of 2, has $(n!)^{n!}$ formal determinants.) Prove or disprove the following two assertions. (1) If M is invertible, then at least one of its formal determinants is invertible. (2) If all the formal determinants of M are invertible, then so is M.

5728. *Proposed by F. N. Fritsch, Lawrence Radiation Laboratory, University of California, Livermore*

Let $X=\{x_1, \cdots, x_N\}$ be a finite subset of E_n with $N\leqq 2n$. Show that there exist two hyperplanes $\mathcal{H}$ and $\mathcal{K}$ such that: (i) $X\subset\mathcal{H}\cup\mathcal{K}$; (ii) $\mathcal{H}$ and $\mathcal{K}$ are not parallel.

5729. *Proposed by L. F. Kemp, Jr., Polytechnic Institute of Brooklyn*

Let $u\times v$ and $u'\times v'$ be two Cartesian product measures defined on $(X\times Y, S\times T)$, the Cartesian product of two measure spaces (X, S) and (Y, T), then

1. $u\times v\ll u'\times v'$ if and only if $u\ll u'$ and $v\ll v'$.
2. $u\times v\perp u'\times v'$ if and only if $u\perp u'$ or $v\perp v'$.

5730. *Proposed by K. D. Juhlin, University of Illinois*

Let E_1^u be the real line with the upper limit topology (basis consists of all sets of the form $(a, b]$). Is $(0, 1]$, with the topology induced by the topology on E_1^u homeomorphic to E_1^u (i.e., is E_1^u homeomorphic to one of its basic open sets?)?

5731. *Proposed by L. N. Childs, State University of New York at Albany*

(a) Let M, N be finite dimensional vector spaces over a field R. Let S be a field extension of R and let M_S, N_S denote the S-spaces generated by M and N ($M_S=M\otimes_R S$, etc.). Prove the following result:

(*) If x in $M\otimes_R N$ has the property that for some S as above, $x\otimes 1=x_S$ in $M_S\otimes_S N_S$ has the form $x_S=y\otimes_S z$, y in M_S, z in N_S, then $x=u\otimes v$, u in M, v in N.

(b) Assume now only that M, N are finitely generated free modules over a commutative ring with unity R, and let S be an R-algebra which is a finitely generated projective R-module. Find conditions on R so that (*) is still true.

5732. *Proposed by James Chew, University of Akron, Ohio*

Prove or disprove: Let $(X, \cdot, \mathfrak{J})$ be a system such that $(X, \cdot)$ is a group and $(X, \mathfrak{J})$ is a topological space such that multiplication is continuous.

If card $(W\cap W^{-1})\geqq 2$ for every open set W containing the identity, then $(X, \cdot, \mathfrak{J})$ is a topological group.

5733. *Proposed by M.F. Neuts, Cornell University*

Let X be a nonnegative integer-valued random variable, and suppose that Y is a random variable satisfying $0\leqq Y\leqq X$. We are interested in the property $(\mathcal{I})$: Y and $Y-X$ are independent.

(A) If Y takes only integer values, and if the conditional distribution of Y, given the value of X, is uniform on $\{0, 1, \cdots, X\}$ show that $(\mathcal{I})$ holds if and only if Y and $Y-X$ have geometric distributions with the same parameter.

(B) If Y takes real values and has uniform conditional distribution, given X on the interval $0\leqq Y\leqq X$, when does $(\mathcal{I})$ hold?

SOLUTIONS OF ADVANCED PROBLEMS

Coverings for Sets of Measure Zero

5665 [1969, 423]. *Proposed by Peter Ungar, New York University*

A student claimed that an arbitrary set of measure zero can, for any $\epsilon > 0$, be covered by a family of intervals $I_1, I_2, \cdots$, such that the length of I_n is $< \epsilon/2^n$. Is this true?

I. *Solution by William Beyer and David Kahaner, Los Alamos Scientific Laboratory.* The claim is false. If such a covering were possible then

$$\lim_{\epsilon \to 0} \sum_{n=1}^{\infty} \left(\frac{\epsilon}{2^n}\right)^p = \begin{cases} 0 & p > 0 \\ \infty & p = 0. \end{cases}$$

Thus the Hausdorff dimension of this set is zero. On the other hand there exist sets of measure zero with positive Hausdorff dimension, e.g., the Cantor set has Hausdorff dimension log 2/log 3. See Hurewicz and Wallman, *Dimension Theory*, p. 102, also Hausdorff, *Dimension und äusseres Mass*, Math. Ann., 79 (1919) 163.

II. *Note by the proposer.* Given a set S of measure zero, the convergent series of positive numbers fall into two classes, depending on whether for every $\epsilon > 0$, S has a covering by a set of intervals $I_1, I_2, \cdots$ with the length of $I_n < \epsilon a_n$, or not. The question arises whether for every set S there exists a convergent series for which coverings of the above kind exist. Given S, such a series can in fact be constructed as follows. Let $I_{k1}, I_{k2}, \cdots$ be a covering of S by intervals of total length $\leqq (\frac{1}{4})^k$, for $k = 1, 2, \cdots$. Enumerate this double infinity of intervals in some manner. For the purpose of the rest of the proof, let $c \times I$ denote the interval with the same center as I, and c times as long as I, where c is a positive number. Let I_n be 2^k times the nth interval in the above enumeration, where k is the index of the family of intervals to which the nth interval belongs. Let a_n be the length of I_n. Then $\sum a_n \leqq 1$. Also, for any fixed h, the set of intervals $2^{-h} \times I_n$ $(h = 1, 2, \cdots)$ will contain the hth family named above as a subset and hence cover S.

Extensive work on this kind of problem is found in Sierpinski, *Hypothèse du Continu*. He cites a paper by E. Borel in Bull. de la Société Math. de France, 47 (1919); also a surprising result of Besicovitch (Acta Matematica, 62 (1934)) who disproves Borel's conjecture that if a set can be covered by a sequence of intervals of lengths $a_1, a_2, \cdots$, where $a_1, a_2, \cdots$ is a completely arbitrary sequence of positive numbers, then the set must be countable.

Also solved by Michel Bousquet, R. J. Buck, M. A. Ettrick, K. M. Garg, D. A. Herrero, Mark Kac, O. P. Lossers (Netherlands), Nicholas Passell, J. T. Rosenbaum, A. C. Segal, and the proposer.

Extension of Kantorovic's Inequality

5666 [1969, 423]. *Proposed by Alexandru Lupaş, Cluj, Romania*

Let Φ be a linear space of real valued functions defined on a set X, assume that $1 \in \Phi$. Let F be a positive linear functional on Φ such that $F(1) = 1$. For

$0<m\leqq f_i\leqq M$, $i=1, 2$, prove the inequality

$$F\left(\frac{1}{mf_1+Mf_2}\right)F\left(\frac{f_1f_2}{mf_1+Mf_2}\right)\leqq\frac{1}{4mM} \tag{1}$$

with equality if and only if $f_1=M, f_2=m$.

Solution by Hiroko Hidaka, Osaka Kyoiku University, Japan. Our proof runs along the line of Rennie [this MONTHLY, 70 (1963), 982] in the proof of the Kantorovic inequality. Since $0<m\leqq f_i\leqq M(i=1,2)$, we have $(f_1-M)(f_2-m)\leqq 0$. Since F is positive, linear and $F(1)=1$, we have an analogue of Rennie's inequality:

$$F\left(\frac{f_1f_2}{mf_1+Mf_2}\right)+F\left(\frac{mM}{mf_1+Mf_2}\right)\leqq 1. \tag{2}$$

Now from the fact that the product of two positive numbers does not exceed $\frac{1}{4}$ if their sum is 1, we get the required inequality (1). Obviously, (1) is an extension of the Kantorovic inequality (put $f_1=f_2$).

Equality in (1) does not imply $f_1=M$ and $f_2=m$ as stated. Counterexample: In the algebra of all 2-dimensional vectors (with the coordinate-wise multiplication), suppose $f_1=(m, M)$, $f_2=(M, m)$ and $F((a, b))=b$.

Also solved by D. A. Zave, and by the proposer.

On Riemann Integrability

5667 [1969, 423]. *Proposed by Ray Glenn, Asheville-Biltmore College*

Prove that a function f which has a finite limit at each point of the closed interval $[a, b]$ is Riemann integrable on $[a, b]$.

Solution by Joseph Horowitz and Alan Shuchat, University of Toledo, Ohio. If f has a finite limit at each point of $[a, b]$, then $f=g+h$, where g is continuous, $h=0$ at all points where h is continuous and $\lim_{x\to x_0} h(x)=0$ for all x_0. By considering positive and negative parts separately, we may assume $h\geqq 0$.

Let $A_n=\{x:h(x)\geqq 1/n\}$. If A_n is infinite, then it has an accumulation point x_0 and $\lim_{x\to x_0} h(x)\geqq 1/n$. Thus each A_n is finite and the discontinuities of h form a set of measure zero. Riemann integrability follows.

Also solved by D. T. Adams, J. W. Andrushkiw, Linda W. Brinn, L. E. Clarke (England), R. A. Christiansen, R. L. Cramer, M. R. Cullen, John Dennis, W. G. Dotson, Jr., M. A. Ettrick, R. S. Fishman, R. V. Fuller, D. W. Hadwin, D. R. Horner, Richard Johnsonbaugh, Emmett Keeler, J. F. Leetch, O. P. Lossers (Netherlands), R. J. Loy, R. K. Mueller, D. E. Myers, P. A. Nickel, Simeon Reich (Israel), A. C. Segal, Jean Spitzer, R. A. Struble, C. Vu-Son, and the proposer.

Notes. Vu-Son observes that the conclusion remains valid if the hypothesis is only that right and left hand limits exist at each point. Leetch points out that the problem also appears as a "Quickie" Q 413, MATHEMATICS MAGAZINE, 40 (1967), p. 232.

THE AMERICAN MATHEMATICAL MONTHLY

(FOUNDED IN 1894 BY BENJAMIN F. FINKEL)

THE OFFICIAL JOURNAL OF

THE MATHEMATICAL ASSOCIATION OF AMERICA, INC.

VOLUME 77 NUMBER 5

CONTENTS

(continued on inside cover)

MAY 1970

NOTICE TO AUTHORS

Articles should be typewritten on good quality paper, triple spaced with wide margins. Submit the original and a duplicate; keep a complete copy for protection against loss. Please follow the format used in current issues of the MONTHLY. The manuscript must make the author's intent clear to the printer. A paper accepted for publication by the editor but not acceptable to the printer will be returned to the author for retyping. Please use the American Mathematical Society *Manual for Authors of Mathematical Papers* as your guide and please conform to correct English usage. A *Manual for Monthly Authors* is available on request.

Prospective MONTHLY authors are advised to consult the Statement of Policy in the January 1969 issue, p. 2. Specialized research is usually unsuitable. Backlog: Main Articles 11 months, Math. Notes 10 months, Research Problems 6 months, Classroom Notes 7 months, Math. Education 6 months.

EDITORIAL CORRESPONDENCE: to HARLEY FLANDERS, American Mathematical Monthly, Purdue University, Lafayette, IN 47907; ADVERTISING CORRESPONDENCE: to RAOUL HAILPERN, Mathematical Association of America, SUNY at Buffalo, N. Y. 14214; CHANGE OF ADDRESS and SUBSCRIPTIONS: to A. B. WILLCOX, Mathematical Association of America, 1225 Connecticut Ave., N.W., Washington, D.C. 20036

Annual dues for members of the Association (including a subscription to the American Mathematical Monthly) are $10.00. For nonmembers the subscription price is $15.00.

PUBLISHED BY THE ASSOCIATION at Washington, D. C., and Menasha, Wisconsin during the months of January, February, March, April, May, June–July, August–September, October, November, December.

Second-class postage paid at Washington, D. C., and additional mailing offices. Acceptance for mailing at special rate of postage provided for in the Act of February 28, 1925, embodied in Paragraph 4, Section 538, P. L. and R., authorized April 1, 1926.

PRINTED IN THE UNITED STATES OF AMERICA

"PROBLEMS IN THE THEORY OF RINGS" REVISITED

IRVING KAPLANSKY, University of Chicago

Introduction. On June 6–8, 1956 the Division of Mathematics of the National Academy of Sciences-National Research Council sponsored a conference on linear algebras, held at Ram's Head Inn on Shelter Island, Long Island, New York. I ventured to give a talk in which I proposed twelve problems in the theory of rings. In due course the proceedings of the conference were published, and [**31**] is the manuscript of my talk. In an introductory paragraph I expressed the hope that these problems would help to rekindle interest in the theory of rings. Whether or not the problems had anything to do with it, it is a pleasant fact that there has since been plenty of activity in ring theory. In particular, six of the twelve problems have been solved, and there are partial results on three others. Incidentally, when Herstein reviewed [**31**] he was already able to report three problems solved and progress on one more.

In this paper I propose to revisit the twelve problems and survey their current status with reasonable completeness. Where new problems suggest themselves, I record them, and I continue the numbering from 13 on. For the reader's convenience, and because [**31**] may not be too readily accessible, the original problems are reproduced.

I have not at all attempted to review all the major progress in ring theory since 1956; I stick rather closely to topics connected with the twelve problems. Particularly notable among things I do not touch are the work of Goldie on noncommutative Noetherian rings, and that of Cohn on free ideal rings and related topics. I hope that some abler pen will provide a comprehensive survey before too long.

I owe Herstein a great debt for his willingness to spend many hours sharing with me his thorough knowledge of the twelve problems, and indeed of all of ring theory. Such merits as this enterprise possesses are virtually entirely due to him. I will venture only to claim full credit for all the boners. Thanks are also due to Amitsur, Bergman, Cohn, Connell, Passman, Rosenberg, and Villamayor for valuable comments on a draft of the paper.

PROBLEM 1. *Is every right primitive ring left primitive?*

The answer is "no," in the light of Bergman's example [**8**]. The example is also presented in [**23**, pp. 122-5], and is sketched on p. 255 of [**26**].

Irving Kaplansky was Saunders Mac Lane's first PhD student. He did his undergraduate work at Toronto and graduate work at Harvard. He was an instructor for several years at Harvard, then in defense work for a year at Columbia. He joined the Chicago faculty after the war, served as department chairman 1962–67, and currently holds the George Herbert Mead Distinguished Service Professorship. He held a Guggenheim Fellowship in 1948–49, and in 1965–66 he spent a leave at Queen Mary College, London.

He was elected to the National Academy of Sciences in 1966, is a trustee of the American Mathematical Society, and received honorary degrees from the University of Waterloo and Queen's University. His book, *Commutative Rings* (Allyn and Bacon), has just appeared. *Editor.*

In [27] Jategaonkar presents further examples which have the additional property of being principal left ideal domains.

It may be useful to investigate whether suitable additional hypotheses (e.g. chain conditions) imply that right and left primitivity are equivalent. With one eye on Jategaonkar's examples, we ask:

Problem 13. If a ring R is right primitive and right and left Noetherian, is R left primitive?

PROBLEM 2. *Does there exist a simple radical ring?*

That the answer is "yes" was announced by Sasiada in [60]. Full details of the argument were furnished by Cohn and Sasiada [10]. See also the account on pp. 125–131 of [23]. Sasiada's example is sketched on pp. 254–5 of [26]. It is worth noting that Sasiada's construction uses power series and consequently yields a ring with cardinal number at least that of the continuum. Hamsher [23, p. 131] and Rjabuhin [56] independently observed that it is possible to drop down to a countable example. Indeed, any countable subset of a simple radical ring can be enlarged to a countable simple radical ring.

Posner [54] encountered simple radical rings in his investigation of noncommutative valuation rings. Osofsky [50] found them pertinent in her work on injective envelopes of cyclic modules. In both cases the hypothetical simple radical rings are not known to exist.

The existence of simple radical rings and the existence of nil algebras which are not locally nilpotent (see Problem 3 immediately following) presumably enhance the chances for an affirmative answer to the next problem.

Problem 14. Does there exist a simple nil ring?

Some properties of simple nil rings (if they exist) are noted by McWorter in [47]. As a final comment on Problem 2, I note Herstein's proof of the nonexistence of simple radical rings satisfying a polynomial identity. This was mentioned in [31]; full details are given by Susan Montgomery in her thesis [49].

PROBLEM 3. *Is every algebraic algebra locally finite?* (*Kurosh's problem.*)

Kurosh proposed the problem in [36]. Golod furnished a negative answer in [16], in the form of a finitely generated nonnilpotent nil algebra. The example was a byproduct of his remarkable joint work with Shafarevich on the class field tower problem [17]. Vinberg [66] added some refinements. Later accounts of the work appear in [15], [23, Chapter 8], and [24, pp. 116–121].

Golod's example simultaneously dragged down to defeat the (unbounded) Burnside problem for groups. It is also reported that various people have constructed in this way an Engel group that is not locally nilpotent. (For background on this problem see [19].)

In view of the Golod example, the affirmative answer [29] to Kurosh's problem for algebras satisfying a polynomial identity seems to stand as a twenty-year-old high water mark of work on the plus side of the ledger. The simplifications introduced by Levitzki [44] should be noted, and excellent accounts appear in [26, Th. 1, p. 242] and [23, pp. 162–168].

Special cases of Kurosh's problem still invite attention. In a draft of this paper I asked whether a primitive algebraic algebra is necessarily locally finite. In a letter dated July 23, 1969, Amitsur described how to "surround" the Golod example by a primitive algebraic algebra which is not locally finite. However, the algebra is not itself finitely generated, so the first half of Problem 15 remains open.

Problem 15. Is every finitely generated primitive algebraic algebra finite-dimensional? Failing this, is every finitely generated algebraic division algebra locally finite? (It is equivalent to ask whether every algebraic division algebra is locally finite.)

The result [**30**, Th. 10.4] that reduced Kurosh's problem to the primitive and nil cases must now be consigned to oblivion as doubly useless. But when a fixed algebra is under consideration, a useful result still survives: If A is an algebraic algebra such that every primitive image of A is locally finite and the radical of every homomorphic image of A is locally finite, then A is locally finite ([**44**, Th. 7.2] or [**26**, Th. 3 on p. 244]; these accounts correct an inaccuracy in [**30**, Th. 10.3]). Still other conditional theorems are [**30**, Th. 10.5] and Corollaries 1 and 2 on p. 409 of [**44**].

PROBLEM 4. *Can any ring with a polynomial identity be embedded in a matrix ring over a commutative ring?*

The answer is "no." Indeed a ring with a polynomial identity need not satisfy any standard identity. Jacobson [**26**, p. 260] attributes the first example to Drazin, and then (as Herstein did in his review of [**31**]) gives P. M. Cohn's example of an infinite-dimensional exterior algebra over any field of characteristic 0. This satisfies $[[xy]z]=0$ but violates the standard identity of any degree.

On the affirmative side, we have Posner's theorem that any prime ring satisfying a polynomial identity can be embedded in a matrix ring over a field ([**53**] or [**23**, Th. 7.3.2] or [**24**, Th. 5.6]). This is a sharpening of earlier results of Amitsur [**1**].

In the first draft of this paper, I asked whether Problem 4 might be salvaged by demanding that the ring satisfy the standard identity in $2n$ variables for some n. This too is false, as Amitsur shows in [**5**]. He constructs an example of a ring R which is even a homomorphic image of a subring of n by n matrices over a commutative ring, yet it cannot be imbedded in m by m matrices over a commutative ring for any m, however large.

PROBLEM 5. *Let A_n denote the n by n total matrix algebra over a field. Does there exist a polynomial which always takes values in the center of A_n without being identically* 0*?*

The problem is stated a little carelessly. The polynomial 1 might be offered as an answer! Or, if this is too bizarre, we could add 1 to, say, the standard identity.

Herstein showed me (he credits the observation to John Thompson) that over a finite field one can construct a polynomial in one variable which sends every n by n matrix to a scalar, and not always the same scalar. This is again unsatisfactory. Let us revamp Problem 5, and give it a new number to avoid confusion. While we are at it let us insert the restriction $n \geqq 3$, for the problem is motivated by the polynomial $(xy-yx)^2$ which works so nicely for $n=2$. (Strictly speaking, $(xy-yx)^2$ should be linearized to meet the conditions of Problem 16.)

Problem 16. Let A_n denote the n by n total matrix algebra over a field, $n \geqq 3$. Does there exist a homogeneous multilinear polynomial (of positive degree) which always takes values in the center of A_n without being identically 0?

For the problems on group rings I shall for simplicity stick to group algebras over a field, although the literature contains numerous results concerning group rings over a ring.

The notation is that G is a group, F a field, and $A(G)$ the group algebra of G over F. In this I follow [**31**], but I acknowledge that it might be better to indicate the dependence on F.

PROBLEM 6. *If G has no elements of finite order, does $A(G)$ have no divisors of* 0*?*

There has been very little progress on Problem 6. It was noted in [**31**] that the answer is affirmative if G can be (linearly) ordered, and therefore also if G is abelian (for it is easy to prove that a torsion-free abelian group can be ordered). In [**7**], Banaschewski showed how to bypass the introduction of an ordering, and also extended the investigation to semi-groups. Lagrange and Rhemtulla [**39**] noted that a one-sided ordering of G suffices (we say, for instance, that G is right-ordered if $a<b$ implies $ac<bc$). In his investigation of right-ordered groups, Conrad [**13**, p. 274] observed that a suitable semi-direct product of two infinite cyclic groups can be right-ordered but not ordered. It is historically interesting that the same group appeared in Levi's paper [**43**], and was presumably the first published example of a torsion-free group that cannot be ordered. Bovdi [**9**] answered Problem 6 affirmatively if G satisfies a certain kind of transfinite solvability.

I learned from George Bergman about a list of 101 problems on rings and modules put out in connection with a symposium at Kishinev on September 3–6, 1968. Problem V.2 of this list asks the following: if G is torsion-free, are scalar multiples of group elements the only units in $A(G)$? Most of the remarks made concerning Problem 6 also apply to this question.

We can ask what group-theoretic property of G corresponds to other ring-theoretic properties of $A(G)$. Quite a number of such results are assembled in [**12**]. Probably the most striking is THEOREM 8: *$A(G)$ is prime if and only if G has no finite normal subgroups $\neq 1$.* A charming companion theorem is Passman's THEOREM III in [**51**]: *$A(G)$ has a non-zero nilpotent ideal if and only if the characteristic of F is $p \neq 0$ and G has a finite normal subgroup with order divisible by p.* Lambek [**40**, pp. 151–165] has a discussion of group algebras which includes these two results.

Perhaps the following question deserves attention:
Problem 17: When is $A(G)$ primitive?

PROBLEM 7. *When is it true that $A(G)$ is semi-simple (in Jacobson's sense)? Is it always true if F has characteristic 0?*

First suppose that F has characteristic 0. At the time of [**31**] the only available result was the proof by Hilbert space methods (due to Rickart) of semi-simplicity over the fields of real and complex numbers. The key idea for new progress was the study by Amitsur in [**3**] of the behavior of the radical of general algebras (not necessarily group algebras) under field extensions.

First, under a separable algebraic extension the radical stays put. (As pointed out by Rosenberg in his review of [**3**], this result is implicit in Theorems V. 8.1 and V. 14.1 of [**26**]. In the revised edition of [**26**], the result appears as Theorem 1 on p. 252.) Second, in a transcendental extension a radical can only arise from a suitable nil ideal of the original algebra. In [**3**] these ideas are combined with the Hilbert space result and methods from [**2**] to get the semi-simplicity of $A(G)$ in case F has infinite transcendence degree over the rationals. In particular, there is victory for uncountable F, a result independently proved by Herstein (unpublished).

In [**4**] Amitsur made a drastic advance and simplification. He first (Lemma 1) deftly proved that for F the field of rational numbers $A(G)$ has no nonzero nil ideals; this is the entering wedge for eliminating the Hilbert space methods. It then followed quickly that $A(G)$ is semi-simple for any F not algebraic over the rational numbers, and that if semi-simplicity of $A(G)$ holds over the rational numbers then it holds for any field of characteristic 0. For later accounts of this work see [**22**, pp. 31–46] and [**26**, pp. 251–254].

It is tantalizing that the subject has not budged since then; the full proof of semi-simplicity seems so near and yet may be so far away.

The literature contains a number of theorems proving semi-simplicity for special classes of groups, e.g., [**4**], [**9**], [**12**], [**18**], [**57**], [**61**], [**64**]. I shall not attempt to survey these results.

We move to the case where F has characteristic p. The situation is highly unsatisfactory in that nobody has proposed a plausible conjecture as to what is needed to make $A(G)$ semi-simple. In the first draft of this paper I naively asked whether the absence of finite normal subgroups with order divisible by p might be sufficient for semi-simplicity. (As noted above, this condition is necessary and sufficient for the absence of nilpotent ideals.) D. S. Passman and A. Chalabi independently showed me that this is false.

Passman [**51**] proved that if G has no elements of order p, then $A(G)$ has no nil ideals and in Theorems IV and VI he went on to corollaries analogous to those obtained by Amitsur for characteristic 0. We can make a bolder conjecture.

Problem 18: If G has no elements of order p (where p is the characteristic of F) is $A(G)$ semi-simple?

If Problem 18 turns out to be false, we can try a stronger hypothesis. Problem 19 is proposed as the second question on page 28 of [**23**].

Problem 19: If G is torsion-free is $A(G)$ semi-simple?

We summarize the situation in characteristic p, as regards semi-simplicity: absence of finite normal subgroups with order divisible by p is necessary, but is not sufficient; absence of elements of order p might be sufficient, but is not at all necessary (cf. [52]).

PROBLEM 8. *It is known that if G is locally finite and has no elements of order p in the case of characteristic p, then $A(G)$ is regular in the sense of von Neumann. Is the converse true?*

With the weakened conclusion that G is torsion, this was proved by M. Auslander [6] by homological methods, by McLaughlin [46] by elementary methods, and by Villamayor [63] again by homological methods. (The first two of these were already reported in [31].) Then Villamayor proved the full result in [65]; he did this by homological methods and on page 948 he also gave a very simple elementary proof.

The fourth problem on page 28 of [23] is pertinent.

Problem 20 (Herstein): If $A(G)$ is algebraic over F is G locally finite?

Problem 20 is very easy in characteristic 0, cf. this MONTHLY, problem 4934, 67(1960), p. 927; solution 68(1961), p. 1015.

An affirmative answer to the next problem would unify Problems 8 and 20.

Problem 21 (Herstein): Suppose that every element in $A(G)$ is either right invertible or a left zero-divisor. Does it follow that G is locally finite?

PROBLEM 9. *It is known that if the characteristic is p and G is a locally finite p-group then $A(G)$ is unit element plus a nil algebra. Is the converse true?*

It is neater to state this in terms of the augmentation ideal Δ (Δ is the set of all $\sum \lambda_i g_i$, $\lambda_i \in F$, $g_i \in G$, with $\sum \lambda_i = 0$). Δ is a two sided ideal of codimension one in $A(G)$. Our problem is: if Δ is nil, prove that G is a locally finite p-group. It is easy to see that G is at any rate a p-group. If we strengthen the hypothesis on Δ to the assumption that Δ is locally nilpotent, then it can be proved that G is locally finite as well. See [45] for these results. But Problem 9 as stated remains open.

Problems 9 and 20 suggest the conjecture that Kurosh's problem, although false in general, is true in suitable circumstances inside a group algebra. We might as well go the whole hog.

Problem 22: Is every algebraic subalgebra of a group algebra locally finite?

If Problem 22 turns out to be too daring, we can retreat to (one-sided or two-sided) ideals, or to nil subalgebras, or to both.

I shall record the two problems on group algebras which are posed on pages 122–123 of [33].

Problem 23: In a group algebra are one-sided inverses two-sided?

Hilbert space methods have proved this to be true in characteristic 0; see [48]. But the characteristic p case remains a mystery.

Problem 24: Let u be an idempotent in a group algebra over a field of char-

acteristic 0, $u\neq 0$, 1. Is the coefficient of the unit element in u rational?

It is known that the coefficient in question is an algebraic number such that it and all its conjugates are real and lie between 0 and 1.

Let me record a theorem which is easily proved by the methods of [**48**]: in the group ring of any group over the integers the only idempotents are 0 and 1. For finite groups this is an old result. Generalizing a result of Swan [**62**, p. 571], Coleman [**11**] extended the finite case to group rings over suitable integral domains.

For several interesting results on idempotents in group algebras see Rudin [**59**].

PROBLEM 10. *Let A be a simple associative ring which is an n by n matrix ring, $n\geqq 3$. Let Z be the center of A, G the multiplicative group of regular elements of A, G' the commutator subgroup of G. Is $G'/(G'\cap Z)$ simple?*

(In [**31**] the first G erroneously is G'.)

One is tempted to call the Artinian case "classical," although the full proof by Dieudonné [**14**] is as recent as 1943. In [**31**] it was noted that Problem 10 had been settled affirmatively in only two further cases: (1) $A=$ the ring of all linear transformations on an infinite-dimensional vector space over a division ring, modulo its unique maximal two-sided ideal. This result is due to Rosenberg [**58**]. (2) $A=$ a factor of type II_1. This requires improving Kadison's topological simplicity [**28**] to algebraic simplicity. A sketch of the proof has now appeared in Appendix IV of [**32**].

The problem had not budged an inch until the thesis of Lanski [**41**]. Lanski has answered Problem 10 affirmatively for certain classes of simple rings. His results cover the case where A is algebraic with an infinite center, and also include Rosenberg's theorem. Since completing his thesis, Lanski has extended his work so as to cover the factor of type II_1 result as well.

PROBLEM 11. *Köthe has determined the structure of algebras satisfying*: (1) *algebraic over an algebraically closed field*, (2) *simple with unit*, (3) *of countable dimension*, (4) *locally simple* (*i.e., any finite subset can be embedded in a finite-dimensional simple subalgebra with the same unit*). *Can a structure theory be developed if* (3) *or* (4) *is dropped or weakened?*

The credit solely to Köthe was a slip. What was being quoted was pages 256–257 of Kurosh's paper [**37**]. Köthe [**35**] had shown (and Kurosh reproved) that such an algebra is an infinite tensor product of total matrix algebras. These can be further decomposed to be of prime size. Then Kurosh proved the number of factors for each p to be invariant (this uniqueness proof does not need countability). So the structure referred to in Problem 11 is this: a cardinal number for every prime.

Of course the problem is vaguely stated. But it seems to be fair to say that there has been no progress (presumably because no one has thought about it).

A reasonable weakening of (4) in Problem 11 is to assume the algebra to be

approximable by finite-dimensional semi-simple subalgebras. For an example of what can happen, see [**38**].

PROBLEM 12. *If D is a division ring, can $D=[DD]$? (Here $[DD]$ is the set of all sums of additive commutators $xy-yx$.)*

Harris [**21**] constructed such a division ring. It is noted in [**31**], with full details in [**21**], that as a consequence the theorem of Wedderburn in [**67**] does not extend to Artinian rings. (The hoped for generalization reads as follows: if every element in a ring R is a sum of nilpotent elements, is R nil?)

Lazerson [**42**] constructed a division ring with the stronger property that every element is a commutator (and not merely a sum of commutators). In fact in his example one of the elements forming the commutator can be fixed. In other words, Lazerson's division ring admits a derivation which is onto.

It is claimed in [**34**, p. 4] that Artinian rings are after all not the natural generalization of finite-dimensional algebras over a field, and that this role ought to be taken by rings which are finitely generated modules over a commutative ring. In this version Wedderburn's Theorem does indeed generalize, as was proved independently by Gruenberg [**20**, p. 295] and Herstein and Small [**25**]. Gruenberg's assumption that the ring is Noetherian can be bypassed by the device of dropping down to a subring finitely generated over the integers. Herstein and Small relied on Posner's generalization [**53**] of Wedderburn's theorem to rings satisfying a polynomial identity. Procesi [**55**] has generalized Posner's result by treating the case where "nilpotent" is replaced by "algebraic."

References

1. S. Amitsur, On rings with identities, J. London Math. Soc., 30 (1965) 464–470.
2. ———, Algebras over infinite fields, Proc. Amer. Math. Soc., 7 (1956) 35–48.
3. ———, The radical of field extensions, Bull. Res. Council Israel Sect. F, (1957–8) 1–10.
4. ———, On the semi-simplicity of group algebras, Michigan Math. J., 6 (1959) 251–3.
5. ———, A non-commutative Hilbert basis theorem and subrings of matrices, preprint, 18 pp.
6. M. Auslander, On regular group rings, Proc. Amer. Math. Soc., 8 (1957) 658–664.
7. B. Banaschewski, On proving the absence of zero-divisors for semi-group rings, Canad. Math. Bull., 4 (1961) 225–231.
8. G. Bergman, A ring primitive on the right but not on the left, Proc. Amer. Math. Soc., 15 (1964) 473–5. Correction on page 1000.
9. A. Bovdi, Group rings of torsion-free groups, (Russian) Sibirski Mat. Z., 1 (1960) 555–8.
10. P. Cohn and E. Sasiada, An example of a simple radical ring, J. of Algebra, 5 (1967) 373–7.
11. D. Coleman, Idempotents in group rings, Proc. Amer. Math. Soc., 17 (1966) 962.
12. I. Connell, On the group ring, Canad. J. Math., 15 (1963) 650–685.
13. P. Conrad, Right-ordered groups, Michigan Math. J., 6 (1959) 267–275.
14. J. Dieudonné, Les déterminants sur un corps non commutatif, Bull. Soc. Math. France, 71 (1943) 27–45.
15. I. Fischer and R. Struik, Nil algebras and periodic groups, this MONTHLY, 75 (1968) 611–623.
16. E. S. Golod, On nil algebras and finitely approximable p-groups, Izv. Akad. Nauk. SSSR Ser. Mat., 28 (1964) 273–6. Amer. Math. Soc. Transl. Ser. 2, 48 (1965) 103–6.
17. ———, and I. R. Shafarevich, On the class field tower, Izv. Akad. Nauk. SSSR Ser. Mat., 28 (1964) 261–272. Amer. Math. Soc. Transl. Ser. 2, 48 (1965) 91–102.

18. J. Green and S. Stonehewer, The radicals of some group algebras, J. of Algebra, 13 (1969) 137–142.

19. K. Gruenberg, Two theorems on Engel groups, Proc. Camb. Philos. Soc., 49 (1953) 377–380.

20. ———, The Engel structure of linear groups, J. of Algebra., 3 (1966) 291–303.

21. B. Harris, Commutators in division rings, Proc. Amer. Math. Soc., 9 (1958) 628–630.

22. I. N. Herstein, Theory of Rings, University of Chicago Math. Lecture Notes, 1961.

23. ———, Noncommutative Rings, Carus Monograph No. 15, MAA, 1968.

24. ———, Topics in Ring Theory, University of Chicago Press, 1969.

25. I. N. Herstein and L. Small, Nil rings satisfying certain chain conditions: an addendum, Canad. J. of Math., 18 (1966) 300–2.

26. N. Jacobson, Structure of rings, Amer. Math. Soc. Coll., vol. 37, rev. ed., 1964.

27. A. V. Jategaonkar, A counterexample in ring theory and homological algebra, J. of Algebra, 12 (1969) 418–440.

28. R. V. Kadison, Infinite general linear groups, Trans. Amer. Math. Soc., 76 (1954) 66–91.

29. I. Kaplansky, Topological representation of algebras II, Trans. Amer. Math. Soc., 68 (1950) 62–75.

30. ———, The structure of certain operator algebras, Trans. Amer. Math. Soc., 70 (1951) 219–255.

31. ———, Problems in the theory of rings, pp. 1–3 of report of a conference on linear algebras. NAS-NRC Publ. 502, Washington, 1957. Review by I. N. Herstein, Math. Rev., 20, 3179.

32. ———, Rings of Operators, Benjamin, New York, 1968.

33. ———, Fields and Rings, Univ. Chicago Press, 1969.

34. ———, Commutative rings, 1968 Jeffery-Williams Lecture, Canadian Mathematical Congress, 1969.

35. G. Köthe, Schiefkörper unendlichen Ranges über dem Zentrum, Math. Ann., 105 (1931) 15–39.

36. A. Kurosh, Ringtheoretische Probleme, die mit dem Burnsideschen Problem über periodische Gruppen in Zusammenhang stehen (Russian with German summary), Izv. Akad. Nauk. SSSR Ser. Mat., 5 (1941) 233–240.

37. ———, Direct decompositions of simple rings, Mat. Sb., 11 (1942) 245–264.

38. ———, On the theory of locally simple and locally central algebras, Mat. Sb., 22 (1948) 443–454.

39. R. Lagrange and A. Rhemtulla, A remark on the group rings of order preserving permutation groups, Canad. Math. Bull., 11 (1968) 679–680.

40. J. Lambek, Lectures on Rings and Modules, Blaisdell, New York, 1966.

41. C. Lanski, The group of units of a simple ring, Chicago Ph.D. thesis, 1969.

42. E. Lazerson, Onto inner derivations in division rings, Bull. Amer. Math. Soc., 67 (1961) 356–8.

43. F. W. Levi, Contributions to the theory of ordered groups, Proc. Indian Acad. Sci. Sect. A, 17 (1943) 199–201.

44. J. Levitzki, On the structure of algebraic algebras and related rings, Trans. Amer. Math. Soc., 74 (1953) 384–409.

45. G. Losey, On group algebras of p-groups, Michigan Math. J., 7 (1960) 237–240.

46. J. McLaughlin, A note on regular group rings, Michigan Math. J., 5 (1958) 127–8.

47. W. McWorter, Some simple properties of simple nil rings, Canad. Math. Bull., 9 (1966) 197–200.

48. Susan Montgomery, Left and right inverses in group algebras, Bull. Amer. Math. Soc., 75 (1969) 539–540.

49. ———, The Lie structure of simple rings with involution of characteristic 2, Chicago Ph.D. thesis, 1969.

50. B. Osofsky, Noncommutative rings whose cyclic modules have cyclic injective pulls, Pacific J. Math., 25 (1968) 331–340.

51. D. S. Passman, Nil ideals in group rings, Michigan Math. J., 9 (1962) 375–384.

52. ———, On the semisimplicity of modular group algebras, Proc. Amer. Math. Soc., 20 (1969) 515–9.

53. E. Posner, Prime rings satisfying a polynomial identity, Proc. Amer. Math. Soc., 11 (1960) 180–4.

54. ———, Left valuation rings and simple radical rings, Trans. Amer. Math. Soc., 107 (1963) 458–465.

55. C. Procesi, The Burnside problem, J. of Algebra, 4 (1966) 421–6.

56. Ju. M. Rjabuhin, On countable simple quasi-regular rings, Mat. Zametki, 4 (1968) 399–403.

57. M. Rosen, The Jacobson radical of a group algebra, Michigan Math. J., 13 (1966) 477–480.

58. A. Rosenberg, The structure of the infinite general linear group, Ann. of Math., 68 (1958) 278–294.

59. W. Rudin, Idempotents in group algebras, Bull. Amer. Math. Soc., 69 (1963) 224–7.

60. E. Sasiada, Solution of the problem of existence of a simple radical ring, Bull. Acad. Polon. Sci. Sér. Sci. Math. Astr. Phys., 9 (1961) 257.

61. S. Stonehewer, Group algebras of some torsion-free groups, J. of Alg., 13 (1969) 143–7.

62. R. Swan, Induced representations and projective modules, Ann. of Math., 71 (1960) 552–578.

63. O. E. Villamayor, On the semisimplicity of group algebras, Proc. Amer. Math. Soc., 9 (1958) 621–7.

64. ———, On the semi-simplicity of group algebras II, Proc. Amer. Math. Soc., 10 (1959) 27–31.

65. ———, On weak dimension of algebras, Pacific J. Math., 9 (1959) 941–951.

66. E. Vinberg, On the theorem concerning the infinite-dimensionality of an associative algebra, Izv. Akad. Nauk. SSSR Ser. Mat., 29 (1965) 209–214.

67. J. H. M. Wedderburn, Note on algebras, Ann. of Math., 38 (1937) 854–6.

THE CHAUVENET SYMPOSIUM

J. C. ABBOTT, U. S. Naval Academy, Annapolis, Maryland

On October seventeenth and eighteenth, 1969, the United States Naval Academy[1] and the Mathematical Association of America[2] sponsored a symposium on "The State of Mathematics Today." The Symposium followed the dedication ceremonies for the Naval Academy's new mathematics building, Chauvenet Hall.

The Association was represented at the dedication by its President, Professor Gail Young, who spoke on William Chauvenet and the Chauvenet Prize. He pointed out that Professor Chauvenet was not only a leading mathematician of his day, but was also one of the founders of the Naval Academy and its first head of the Mathematics Department. In recognition of his devotion to the highest ideals of good exposition and teaching, the Association established the Chauvenet Prize in 1925 for outstanding expository papers in mathematics in English. Since then seventeen awards have been made, to mathematicians

[1] The Office of Naval Research supported the Symposium under contract #043-392.

[2] The Symposium replaced the regular Fall Meeting of the Maryland-District of Columbia-Virginia Section of the Association.

whose names have become household words in the mathematical community. Six of these Chauvenet Prize winners participated in the Symposium. The MONTHLY will publish the written versions of the talks given at the Symposium. The first two are included in this issue; three others will appear in the June–July issue; the sixth by Professor Leon Henkin, which had to be postponed due to illness in his family, will appear later.

We have felt that one of the major problems which faces mathematics today is the problem of proliferation. Mathematics is expanding so rapidly that it is no longer possible for any one individual to follow the great traditions of the nineteenth century and encompass any but a small corner in one of the many vast fields of mathematics. Great names such as Euler, Cauchy, Hilbert, *et al.*, who were able to make significant contributions in all areas of mathematics, are much scarcer today. Therefore the objective of this Symposium was to bring together a number of mathematicians from widely diverse areas who are renowned, not only for their research contributions, but also for their ability to communicate their achievements to others not in their special fields. In this way we hoped to make available to the general mathematical public an understanding of the role and importance of the various parts of mathematics and thereby continue the expository tradition set by Professor Chauvenet that "Expository mathematics is good mathematics and good mathematics lends itself to good exposition."

The first paper in this issue is by Professor Paul Halmos of the University of Indiana on "Finite-Dimensional Hilbert Spaces." Like many of this country's most famous mathematicians, Professor Halmos was a product of central Europe having been born in Budapest, Hungary. He came to this country in 1933 and received most of his professional training at the University of Illinois. He had a varied career beginning at Syracuse and including many years at Chicago and Michigan with a recent tour at the University of Hawaii. His well-known texts on Measure Theory, Finite Dimensional Vector Spaces, Naive Set Theory, Boolean Algebra and Algebraic Logic, show the breadth of his interest and have been a constant source of pleasure to many. He has a personal style that few mathematicians achieve and he has always been devoted to excellence in exposition. He received the Chauvenet Prize in 1947 for his paper on "The Foundations of Probability." In his present paper he returns to some unsolved problems in finite dimensional Hilbert spaces using some techniques of general operator theory and lattice theory.

Professor Guido Weiss is the youngest of the Symposium speakers, having received his doctorate from Chicago in 1956. He was born in Trieste, Italy in 1928, and spent some years at DePaul before going to the University of Washington at St. Louis in 1961. His main interests have always been in the area of harmonic analysis and harmonic functions. He received the Chauvenet Prize in 1967 for his paper on "Harmonic Analysis." In the present paper he examines some methods of complex variables in harmonic analysis.

The first paper of the Symposium was to have been by Professor Gordon T.

Whyburn of the University of Virginia. Unfortunately, just before the Symposium the mathematical community was shocked to hear of Professor Whyburn's sudden death. However, following the best traditions of all great teachers, Professor Whyburn had so well prepared his talk that it was nearly completed at the time of his death. Consequently his colleague, Professor Edwin Floyd, also of the University of Virginia, was able to summarize it as a memorial-tribute to Professor Whyburn at the Symposium. Professor Floyd was also able to complete the original Whyburn manuscript so that it will appear as the first article in the June–July issue of this MONTHLY. The article is particularly appropriate since it summarizes Professor Whyburn's work of the last fifteen years on analytic topology.

Professor Whyburn was a product of R. L. Moore at Texas and followed the traditions of this school by developing his own school at the University of Virginia. It has produced some of the world's great topologists, including Floyd himself, J. L. Kelley, R. H. Bing, and Victor Klee amongst many others. His ability as an expositor was recognized early in 1938 when he won the Chauvenet Prize for his paper "On Structure of Continua." He published continuously until his death, advancing the field of analytic topology to its present highly developed state.

The second paper to appear in the June–July issue will be by Professor Saunders MacLane of the University of Chicago. Professor MacLane received his advanced training at Yale and Göttingen and spent his early teaching years at Harvard before transferring to Chicago in 1947. His early interests were classical algebra and the extension problem for groups. This led naturally into his work with Eilenberg in the development of homological algebras, which eventually grew into category theory, one of the fastest growing branches of modern mathematics. He won the Chauvenet Prize in 1941 for his papers on "Modular Fields" and "Some Recent Advances in Algebra." But despite a lifetime in algebra, Professor MacLane has always been proud of a wide interest in all fields of mathematics. Category theory spans many facets. His present paper entitled "Hamiltonian Mechanics" testifies to his versatility in bringing new methods to bear on classical problems.

Of all the Chauvenet Prize winners Professor Mark Kac, of Rockefeller University, is the only one who has been awarded the prize twice. He won the Prize first in 1950 for his paper on "Random Walk and the Theory of Brownian Motion" and again in 1968 for his paper "Can One Hear the Shape of a Drum?" Professor Kac was also of European origin, having been born in Poland and educated at the John Casimir University where he received his doctorate in 1937. His publications range widely throughout the fields of analysis, probability theory, statistics and number theory. His present paper applies recent developments in analysis to probability theory.

All six of the Chauvenet Memorial Lectures will be available bound separately for library and personal use next fall after the appearance of the Henkin lecture. At that time, copies can be ordered from Professor J. C. Abbott, U. S. Naval Academy, Mathematics Department, Annapolis, MD 21402.

FINITE-DIMENSIONAL HILBERT SPACES

P. R. HALMOS, Indiana University

Prologue. I used to like linear algebra because it gave me a motivation for the study of operators on Hilbert space and because it gave me insight into the algebraic skeleton of operator theory, which made that study easier. Now I like what I learned about Hilbert space because it keeps shedding light on more and more new aspects of linear algebra and because it succeeds in keeping that classical subject alive and exciting. The purpose of this report is to illustrate the latter point by describing three non-trivial parts of finite-dimensional linear algebra, the original impetus for which came from operator theory on infinite-dimensional Hilbert space. The subjects are (1) an algebraic characterization of pairs of subspaces of a finite-dimensional Hilbert space, (2) a geometric characterization of linear transformations in terms of rotations and projections (dilation theory), and (3) a statement of some fragmentary results and challenging open problems about lattices of invariant subspaces.

A finite-dimensional Hilbert space is, by definition, a finite-dimensional unitary space (complex inner product space). The only prerequisite for an intelligent reading of this paper is acquaintance with the language, notation, and principal facts of finite-dimensional unitary geometry; see, for instance, [**6**]. As for the insistence on complex numbers: the geometric language of Hilbert space is motivated by the real case, but the algebraic hurdles are most easily overcome if complex numbers are allowed. The customary way out (followed here) is to use complex coefficients, and, at the same time, continue to use real language; this does not seem to lead to serious or permanent confusion.

Two subspaces. What are all the different ways in which a subspace can be placed in a finite-dimensional Hilbert space? The question is vague, but it has a reasonably definite answer. If $H=C^n$ (where C is the complex number field), then one way to get an r-dimensional subspace of H $(0\leqq r\leqq n)$ is to form the set of all those vectors whose last $n-r$ coordinates vanish. More to the point is that to within isomorphism this is the only way; every r-dimensional subspace of C^n can be obtained from this one by a suitable rotation.

What are all the different ways in which two subspaces, or three, or any number can be placed in a finite-dimensional Hilbert space? The difficulty of the answer seems to increase with the number. The preceding paragraph shows that there is no difficulty about putting *one* subspace M into a space H; just put it down, anywhere, let $M^\perp$ fall where it may, and there is nothing left to ask. Before the position of *two* subspaces M and N in H can be said to be known, many questions must be asked and answered. Is M included in N? Is M orthogonal to N? Does $M^\perp$ have a nontrivial intersection with $N^\perp$? If the relation between M and N is not describable in the simple terms of inclusion and orthogonality, does it make sense to ask for the "angle" between them? Such questions were first raised by Dixmier [**3**]; the point of view described below is somewhat dif-

ferent and more recent [**11**]. As for three subspaces, or more, the mind boggles. There is, in fact, reason to believe that the problem of three subspaces will be out of human reach for a long time to come. A comment of Chandler Davis [**2**] indicates that if we knew all about three subspaces, then we could learn more about unitary equivalence than, apparently, we are meant to know.

In the study of pairs of subspaces there are four thoroughly uninteresting cases, the ones in which both M and N are either 0 or H. In the most general case the entire space is the direct sum of five subspaces:

$$M \cap N,\ M \cap N^{\perp},\ M^{\perp} \cap N,\ M^{\perp} \cap N^{\perp},$$

and the rest. The parts of M and N in the first four are "thoroughly uninteresting". In "the rest", the orthogonal complement of the span of the first four, M and N are in *generic position*, in the sense that all four of the special intersections listed above are equal to 0.

The simplest example of two subspaces in generic position consists of two distinct non-orthogonal lines in a plane, and there is no loss of generality in taking one of them to be the first coordinate axis. To get a useful generalization, suppose that T is a non-singular linear transformation on a finite-dimensional Hilbert space K, write $H=K\oplus K$, let M be the "horizontal axis" consisting of all vectors of the form $\langle f, 0\rangle$ in H, and let N be the graph of T, i.e., the set of all vectors of the form $\langle f, Tf\rangle$ in H. The assertion that M and N are in generic position needs a little proof. The first step is to show that $M\cap N=0$. Indeed, how can an $\langle f, 0\rangle$ be equal to a $\langle g, Tg\rangle$? Answer: only if $Tg=0$, whence $g=0$ (because T is non-singular), and therefore $f=0$. For the rest of the proof it is necessary to know $M^{\perp}$ (trivial: all $\langle 0, f\rangle$) and $N^{\perp}$ (easy and standard computation: all $\langle -T^*f, f\rangle$). From this it is easy to deduce that $M^{\perp}\cap N^{\perp}=0$: since T^* is just as non-singular as T, the proof just given applies again. The equations $M\cap N^{\perp}=0$ and $M^{\perp}\cap N=0$ are trivial.

The basic result in the theory of two subspaces is that this way of constructing pairs of subspaces in generic position is, to within unitary equivalence, the only way. More precisely: *if M and N are subspaces in generic position in a finite-dimensional Hilbert space H, then there exists a finite-dimensional Hilbert space K, and there exists a non-singular linear transformation T on K, such that the pair $\langle M, N\rangle$ is unitarily equivalent to the pair $\langle K\oplus 0$, graph $T\rangle$.*

What follows is an outline of the proof; with suitable analytic caution the proof is generalizable to the infinite-dimensional case. Let P be the projection with range M. Assertion: the restriction of P to N is a non-singular linear transformation from N onto M. Suppose, indeed, that $Pg=0$ for some g in N. It follows that $g\in M^{\perp}\cap N$, and hence (generic position) that $g=0$; this proves that the kernel of P in N is 0. To prove that the image PN is equal to M, suppose that $f\in M$ and $f\perp PN$. This means that if $g\in N$, then $0=(f, Pg)=(Pf, g)=(f, g)$, so that $f\in M\cap N^{\perp}$. It follows (generic position) that $f=0$; the proof of the assertion is complete.

The existence of a non-singular linear transformation from N onto M implies that M and N have the same dimension. (This could have been proved more quickly, but the slower approach is needed for the rest of the proof anyway.) Since M and $M^{\perp}$ on the one hand and N and $N^{\perp}$ on the other hand enter the hypotheses with perfect symmetry, it follows that all four of these subspaces have the same dimension. Since this applies to M and $M^{\perp}$ in particular, there exists an isometric linear mapping from M onto $M^{\perp}$; the idea from now on is to identify each element of $M^{\perp}$ with the element of M that it corresponds to.

Now put $K=M$. To define T at an element f of K, recall first that $f=Pg$ for a uniquely determined vector g in N, project g into $M^{\perp}$, and let Tf be the element of M that is identified with the element of M so obtained. (A simple 2-dimensional picture should make that long-winded sentence crystal clear.) The verification that K and T do what is expected of them is straightforward.

Given a line in the plane, distinct from both the horizontal and the vertical axes, rotate the plane through the negative of half the angle of inclination. The given line and the horizontal axis become, after the rotation, a line and its reflection through the (new) horizontal axis. This half-angle rotation can be generalized to yield a different and useful representation for a pair of subspaces in generic position; the result is that any such pair is unitarily equivalent to a pair of the form $\langle\text{graph } T_0, \text{graph } (-T_0)\rangle$, for a suitable linear transformation T_0. From this representation, in turn, it is easy to recapture Dixmier's main theorem on pairs of subspaces in generic position: the result is that a *single* Hermitian transformation, namely the sum of the two projections whose ranges are the given subspaces, constitutes a complete set of unitary invariants for the pair.

Unitary dilations. Suppose that H is a subspace of a finite-dimensional Hilbert space K, and let P be the projection from K onto H. Each linear transformation B on K induces in a natural way a linear transformation A on H defined for each f in H by

$$Af = PBf.$$

Under these conditions the transformation A is called the *compression* of B to H, and B is called a *dilation* of A to K. This geometric definition of compression and dilation is to be contrasted with the customary concepts of restriction and extension: if it happens that H is invariant under B, then it is not necessary to project Bf back into H (it is already there), and, in that case, A is the restriction of B to H, and B is an extension of A to K. Restriction-extension is a special case of compression-dilation, the special case in which the linear transformation on the larger space leaves the smaller space invariant.

Compressions and dilations can be usefully described in terms of matrices. If K is decomposed into H and $H^{\perp}$, and, correspondingly, transformations on K are written in terms of matrices (whose entries are transformations on H, and on $H^{\perp}$, and between the two), then a necessary and sufficient condition that B be

a dilation of A is that the matrix of B have the form

$$\begin{pmatrix} A & X \\ Y & Z \end{pmatrix}.$$

The purpose of dilation theory is to get information about difficult transformations by finding their easy dilations. The program is spectacularly successful. Unitary transformations (rotations) are among the easiest to deal with, and it turns out that, except for an easily adjusted normalization, every transformation has a unitary dilation. Some normalization is clearly necessary: if B is unitary, then $\|Bf\| = \|f\|$ for every vector f, and it follows that $\|Af\| \leqq \|f\|$ for every vector f; in other words, if A has a unitary dilation, then A must not increase the norm of any vector. (In the appropriate geometrical technical term, A must be a *contraction.*) That much normalization is sufficient: *every contraction has a unitary dilation.*

As a heuristic guide to the proof, consider the very special case in which the given Hilbert space is 1-dimensional real Euclidean space and the dilation space K is the plane. In that case the given contraction is a scalar α (with $|\alpha| \leqq 1$), and, in geometric terms, the assertion is that multiplication by α (on the line) can be achieved by a suitable rotation (in the plane), followed by projection (back to the line). A picture makes all this crystal clear again.

The proof in the general case can be obtained by first transcribing the synthetic proof just outlined to analytic form, and then imitating the analytic geometry with matrices in the place of numbers. The conceptual problems that the program encounters are familiar ones, and so are their solutions; see [5] for the details.

The least unitary looking contraction is 0, but, of course, even it has a unitary dilation; one such is

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

This dilation does not have many useful algebraic properties. It is not necessarily true, for instance, that the square of a dilation is a dilation of the square; indeed, the square of the dilation of 0 exhibited above is

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

which is not a dilation of the square of 0. Is there a unitary dilation of 0 that is fair to squares? The answer is yes:

$$\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

is an example. The square of this dilation is

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix},$$

which is a dilation of the square of 0. Unfortunately, however, this dilation is not perfect either; its cube is

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

which is not a dilation of the cube of 0. That is (in self-explanatory language): the 3×3 matrix is a 2-dilation of 0, but not a 3-dilation.

Question: does every contraction on an n-dimensional Hilbert space have a unitary k-dilation for every k? Answer: yes, on a Hilbert space of dimension $n(k+1)$; an elementary proof was given by Egerváry [**4**]. More is true: Nagy proved [**15**] that every contraction A has a unitary dilation B such that B^k is a dilation of A^k simultaneously for every positive integer k. (Nagy's paper came between [**5**] and [**4**].) The result is true for infinite-dimensional Hilbert spaces too, but its impressive generality has a price: even if the given space H is finite-dimensional, the dilation space K may have to be infinite-dimensional.

To show how dilation theory can be used, consider the finite-dimensional special case of a beautiful and powerful analytic theorem of von Neumann [**18**]. The assertion is that if A is a contraction and if q is a polynomial such that $|q(z)|\leqq 1$ whenever $|z|=1$, then $q(A)$ is a contraction. The proof via dilation theory goes as follows: If the degree of q is k, find a unitary k-dilation B of A. It follows then that $q(B)$ is a dilation of $q(A)$, and hence it is sufficient to prove that $q(B)$ is a contraction. (In other words, dilation theory reduces the problem to the consideration of unitary transformations only.) But that is trivial: a unitary B has a diagonal matrix, whose diagonal entries are complex numbers of modulus 1; since the corresponding matrix of $q(B)$ has diagonal entries whose moduli are not greater than 1, the desired conclusion becomes obvious.

Reflexive lattices. The set of all subspaces of a finite-dimensional Hilbert space H is a lattice (with respect to the operations of intersection and span) with zero element 0 and unit element H. Certain of its sublattices, the ones called *reflexive*, are of interest in linear algebra. The definition of reflexivity requires of a lattice $\mathfrak{L}$ that a two-step process performed on $\mathfrak{L}$, which always yields a lattice of subspaces at least as large as $\mathfrak{L}$, should, in fact, yield exactly $\mathfrak{L}$, and nothing more. The two steps are these: (1) form all linear transformations that leave invariant each subspace of $\mathfrak{L}$, and then (2) form all subspaces of H that are invariant under all those linear transformations.

Here is an example. Suppose that H is 2-dimensional, and let $\mathcal{L}$ consist of 0, H, and two distinct lines (i.e., 1-dimensional subspaces of H). To say of a linear transformation that it leaves invariant each subspace in $\mathcal{L}$ is to say just that it has two prescribed eigenvectors, and hence that its matrix with respect to the basis they form is diagonal. Since the only subspaces simultaneously invariant under all such diagonal transformations are the ones in $\mathcal{L}$, the lattice $\mathcal{L}$ is reflexive indeed.

Here is a non-example. Suppose again that H is 2-dimensional, and let $\mathcal{L}$ consist of 0, H, and three distinct lines. It is very hard for a linear transformation on H to have three distinct eigenvectors; the only linear transformations that can do it are the scalar multiples of the identity. Such scalar multiples, on the other hand, leave invariant every subspace of H. The two-step process applied to this $\mathcal{L}$ drastically enlarges $\mathcal{L}$; instead of the 5-element lattice $\mathcal{L}$, the enlargement is the infinite lattice of all subspaces of H.

The non-example of the preceding paragraph fails to be reflexive the worst way anything can; the enlargement it effects is maximal. Another way of saying the same thing is that a linear transformation that leaves invariant every subspace of the lattice is necessarily a scalar. A lattice that is non-reflexive in this extreme way is called *transitive.*

A basic open problem about operator theory on infinite-dimensional Hilbert spaces is to characterize all reflexive lattices and all transitive (extremally non-reflexive) lattices. A little progress has been made, but not very much. The finite-dimensional specialization of what is known amounts to two statements: (1) every chain (totally ordered set) of subspaces is reflexive, and (2) every Boolean algebra of subspaces is reflexive. The infinite-dimensional case of (1) is Ringrose's generalization [**17**] of a result of Kadison and Singer [**14**]. For a statement of the appropriate infinite-dimensional formulation of (2) see [**10**]; the proof has not been published yet. In the finite-dimensional case both results become almost trivial.

As for transitive lattices, even less is known. The example above (the one with three lines) is in a certain sense degenerate. From the point of view of projective geometry, which is quite appropriate here, the space of that example has dimension 1, not 2, and the example does not help to answer the question whether higher-dimensional examples exist at all.

An interesting unpublished observation of J. E. McLaughlin shows that they do exist; one such, in C^n, consists of all those subspaces that are invariant under the formation of complex conjugates. More explicitly: call a subspace M of C^n *symmetric* in case M contains, along with each of its vectors, the vector whose coordinates are obtained from the given one by complex conjugation. Assertion: the set of all symmetric subspaces is a transitive lattice. The proof requires a moment's thought, but there is nothing profound about it.

The example of the preceding paragraph yields many examples. Given a finite-dimensional Hilbert space, coordinatize it (i.e., establish an isomorphism

between it and C^n); the lattice of symmetric subspaces with respect to that coordinatization is a transitive lattice. Question: are all transitive lattices obtainable in this way?

The answer is no for two reasons, both trivial, and the heart of the question is still unanswered. *First*, a topological distinction arises: are the lattices under discussion closed or not? (There is only one reasonable topology for the space of subspaces; the question makes unambiguous sense.) If a lattice is dense in a transitive lattice, then it itself is transitive; the question loses no vigor at all if attention is restricted to closed lattices only. As long as restrictions are in order, here is one spot where the complex field makes life more complicated, not less; the question retains all its interest if attention is restricted to real spaces only. *Second*, for spaces of even dimension $2n$ there is a construction that yields a transitive lattice whose non-trivial elements are all of dimension n. Such a lattice imitates an already observed misbehavior; it is isomorphic to a sublattice of the lattice of the projective line.

The result of the indicated specializations is the following question: is every closed transitive lattice of subspaces of an odd-dimensional Euclidean space (real inner product space) equal to the lattice of all subspaces? The answer does not seem to be known.

Note added in proof. K. J Harrison (Monash University) has recently discovered a new transitive lattice of 18 elements that shows that the answer to the question as it stands is no. A modification, however, restores the question: just add the hypothesis that the atoms of the lattice span the whole space. Harrison's discovery makes the problem of determining all transitive lattices even more challenging than before.

Epilogue. Three topics were discussed above to illustrate the thesis that operator theory on Hilbert space yields non-trivial questions and answers about finite matrices. Choosing the examples was not an easy task; many more are available than can be included in one lecture, or one paper, of reasonable length.

I could have chosen the theorem about "near" projections (if the projections onto two subspaces are near enough in norm, then the subspaces have the same dimension) [**8**, Problem 43]. I could have discussed the "power inequality" ($w(A^n) \leqq (w(A))^n$) for the numerical range [**8**, Problem 176], which started in the theory of partial differential equations and ended as a problem about matrices, a problem that refused to become trivial even in the 2×2 case. The topological properties of sets of reducible and irreducible operators were of recent research interest [**9**, **16**], and so also were partial isometries [**7**, **12**]; in both these cases new and interesting facts about the finite-dimensional case emerged. The theory of matrices whose entries are matrices still has some life in it; thus, for instance, the "ultra-invariant" subspaces of binormal operators (2×2 matrices whose entries are commutative normal operators; see [**1**]) are still being studied (by R. G. Douglas and C. M. Pearcy), and there are small,

amusing, and until recently unnoticed questions even about determinants. (Sample: if all four ways of forming the formal determinant of a 2×2 matrix whose entries are matrices yield an invertible matrix, does the matrix itself have to be invertible? See [**13**].)

The subjects just given honorable mention, as well as the three actually discussed in detail, have been receiving serious research attention in the course of the last twenty years (and many still are), and they have all led to questions that could and should have been asked in the finite-dimensional case long before, but as a matter of historical fact they were not. The reason perhaps is that the powerful tools of finite-dimensional linear algebra are too good; they sometimes conceal the elegant and intricate structure that the difficulties of the infinite-dimensional theory bring out. Most of the problems of operator theory can be formulated in the finite-dimensional case, and there are two reasons why it is good to do so. The old reason is that the finite can suggest what should and should not be tried with the infinite; the new reason is the joy of seeing the infinite inspire and guide the finite and contribute to a new flowering of an old subject.

References

1. A. Brown, The unitary equivalence of binormal operators, Amer. J. Math., 76 (1954) 414–434.
2. C. Davis, Generators of the ring of bounded operators, Proc. Amer. Math. Soc., 6 (1955) 970–972.
3. J. Dixmier, Position relative de deux variétés linéaires fermées dans un espace de Hilbert, Rev. Sci., 86 (1948) 387–399.
4. E. Egerváry, On the contractive linear transformations of n-dimensional vector space, Acta Sci. Math. (Szeged) 15 (1953) 178–182.
5. P. R. Halmos, Normal dilations and extensions of operators, Summa Brasil. Math., 2 (1950) 125–134.
6. ———, Finite-dimensional Vector Spaces, Van Nostrand, Princeton, 1958.
7. ——— and J. E. McLaughlin, Partial isometries, Pacific J. Math., 13 (1963) 585–596.
8. ———, A Hilbert Space Problem Book, Van Nostrand, Princeton, 1967.
9. ———, Irreducible operators, Michigan Math. J., 15 (1968) 215–223.
10. ———, Invariant subspaces, Abstract spaces and approximation, Proc. MRI Oberwolbach (1968).
11. ———, Two subspaces, Trans. Amer. Math. Soc., 144(1969) 381–389.
12. ——— and L. J. Wallen, Powers of partial isometries, J. Math. Mech., 19 (1970) 657–663.
13. ———, Advanced problem 5727, this MONTHLY, 77 (1970) 409.
14. R. V. Kadison and I. M. Singer, Triangular operator algebras, Amer. J. Math., 82 (1960) 227–259.
15. B. Sz.-Nagy, Sur les contractions de l'espace de Hilbert, Acta Sci. Math. (Szeged) 15 (1953) 87–92.
16. H. Radjavi and P. Rosenthal, The set of irreducible operators is dense, Proc. Amer. Math. Soc., 21 (1969) 256.
17. J. R. Ringrose, On some algebras of operators, Proc. London Math. Soc., 15 (1965) 61–83.
18. J. von Neumann, Eine Spektraltheorie für allgemeine Operatoren eines unitären Raumes. Math. Nachr., 4 (1951) 258–281.

COMPLEX METHODS IN HARMONIC ANALYSIS

GUIDO WEISS, Washington University, St. Louis

1. Introduction. The purpose of this paper is two-fold. First, we plan to show how properties of analytic functions of a complex variable can be used to obtain several results of classical harmonic analysis (that is, the theory of Fourier series and integrals of one real variable). This will be done in Section 1. Second, in Section 2 we shall indicate how some of these applications of the theory of functions can be extended to Fourier analysis of functions of several variables.

Some of these applications of the theory of functions seem very startling since the results obtained appear to involve only the theory of functions of a *real* variable or the theory of measure. This is true of the following results all of which were originally obtained by these complex methods. In fact, "real variable" proofs of (a) and (c) are still not known.

(a) Suppose μ is a finite Borel measure on $(-\pi, \pi)$, then its *Fourier-Stieltjes* coefficients are $c_n = (1/2\pi)\int_{-\pi}^{\pi} e^{-in\theta}\, d\mu\,(\theta)$, $n = 0, \pm 1, \pm 2, \cdots$. If $c_n = 0$ for $n < 0$ then μ is absolutely continuous. This result, obtained by F. and M. Riesz [9], has many applications. One of them is the following result of Helson [8]: If $s_n(\theta) = \sum_{k=-n}^{n} c_k e^{ik\theta}$, $n = 1, 2, 3, \cdots$, are the partial sums of a trigonometric series and they satisfy

$$\|s_n\|_1 = \int_{-\pi}^{\pi} |\, s_n(\theta)\,|\, d\theta \leqq A < \infty,$$

where A is independent of $n = 1, 2, 3, \cdots$, then the coefficients c_n tend to 0. Since $\int_{-\pi}^{\pi} e^{in\theta} d\theta = 0$ when $n \neq 0$ this applies to the important special case $s_n(\theta) \geqq 0$ for $n = 1, 2, 3, \cdots$.

(b) If $f \in L^p(-\infty, \infty)$, $1 \leqq p < \infty$, then the *Hilbert transform* $\tilde{f}$ of f is defined by the limit

$$\tilde{f}(x) = \lim_{\epsilon \to 0} \left\{ \frac{1}{\pi} \int_{|t| \geqq \epsilon > 0} f(x-t)\, \frac{dt}{t} \right\}. \tag{1.1}$$

This definition, of course, only makes sense if the limit on the right exists. It can be shown that this limit exists and is finite a.e. If we also assume that $p \neq 1$,

$$\|\tilde{f}\|_p \leqq A_p \|f\|_p,$$

where A_p is independent of $f \in L^p$ (this result is known as the *M. Riesz inequality*). When $p = 1$ it is not true that $\tilde{f}$ belongs to L^1; however, the following more general condition is satisfied by $\tilde{f}$: if *m denotes Lebesgue measure then*

$$m(\{x \in (-\infty, \infty) : |\, \tilde{f}(x)\,| > s \geqq 0\}) \leqq \frac{A}{s} \|f\|_1, \tag{1.2}$$

where A is independent of $f \in L^1(-\infty, \infty)$. (Inequality (1.2) is a special case of

what are known as *weak-type* inequalities for operators on L^p spaces. For the definition of this notion and for many illustrations of its use, we refer the reader to Chapter XII of [**16**].)

(c) A recently obtained theorem of Calderón [**3**], having important applications in the theory of partial differential equations, extends the last result: Suppose that a is a Lipschitz function on $(-\infty, \infty)$, then the operator

$$T_a(f)(x) = \lim_{\epsilon\to 0} \int_{|x-t|\geq\epsilon>0} f(t) \frac{a(x) - a(t)}{(x-t)^2}\, dt$$

is well defined almost everywhere when $f\in L^p(-\infty, \infty)$, $1<p<\infty$, and satisfies the inequality $\|T_af\|_p \leqq A_p\|f\|_p$, where A_p is independent of f.

(d) The Fourier series analogue of the Hilbert transform is the conjugate function operator which can be defined for $f\in L^1(-\pi, \pi)$ by the principal value integral

$$\tilde{f}(\theta) = \lim_{\epsilon\to 0} \frac{1}{\pi} \int_{\pi\geq|\phi|\geq\epsilon>0} f(\theta - \phi) \frac{\sin\phi}{1-\cos\phi}\, d\phi.$$

This operator satisfies the properties described in example (b). That is, $\tilde{f}(\theta)$ exists almost everywhere and, if $1<p<\infty$, there is a constant A_p, independent of f, such that $\|\tilde{f}\|_p \leqq A_p\|f\|_p$.

The classical theory of Fourier series has been described by A. Zygmund, in the preface of his book [**16**], "As the meeting ground of real and complex variables." This should come as no surprise once we make the following observations. Suppose

$$\frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos k\theta + b_k \sin k\theta) \tag{1.3}$$

is a trigonometric series with real coefficients. Then if we add to it (formally) the series

$$i\sum_{k=1}^{\infty} (a_k \sin k\theta - b_k \cos k\theta)$$

we obtain

$$\frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k - ib_k)e^{ik\theta},$$

which is the restriction to the unit circle $|z| = 1$ of the power series

$$\frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k - ib_k)z^k. \tag{1.4}$$

We see, therefore, that each trigonometric series with real coefficients is, formally, the real part of a power series development about the origin. Hence, it is not

unreasonable to expect that some results concerning trigonometric and Fourier series could be obtained by the following general procedure. If we desire to establish some properties of the series (1.3) we construct the power series in the manner described above, which, hopefully, defines an analytic function F in the interior of the unit circle. We may, if this is the case, then apply results of the theory of functions to obtain certain properties of F. By tending to the boundary of the unit disc in a radial direction (that is, by having $z=re^{i\theta}$ approach $e^{i\theta}$ by letting r tend to 1) we may then expect to obtain an almost everywhere defined function of θ having properties enabling us to obtain the desired result. In case we are dealing with functions defined on the entire real line and their Fourier integrals the situation is much the same. In this case we might think of the real line as embedded in the plane with the upper half plane assuming the role of the interior of the unit circle.

A classical theorem of Fatou [6] is, perhaps, the most basic tool that can be used for carrying out the above described procedure. This theorem asserts that any bounded analytic function $F(z)=F(re^{i\theta})$ defined for $|z|<1$ has radial limits almost everywhere. That is, for almost every $\theta\in(-\pi,\pi)$, $\lim_{r\to 1} F(re^{i\theta})=F(e^{i\theta})$ exists. The corresponding result for a bounded analytic function F in the upper half plane $\{z=x+iy:\ -\infty<x<\infty,\ 0<y\}$ is that $\lim_{y\to 0} F(x+iy)=F(x)$ exists for almost every x.

Let us describe in more detail how Fatou's theorem can be applied to obtain result (b) stated above. Suppose, then, that $f\in L^p(-\infty,\ \infty)$, $1\leqq p<\infty$. We want to show that

$$\int_{|t|\geqq\epsilon>0} f(x-t)\,\frac{dt}{t}$$

tends to a limit as ϵ tends to 0. This problem is reduced to one involving an analytic function in the following way: We form the Poisson integral

$$u(x,y)=\frac{1}{\pi}\int_{-\infty}^{\infty} f(x-t)\,\frac{y}{t^2+y^2}\,dt$$

and the conjugate Poisson integral

$$v(x,y)=\frac{1}{\pi}\int_{-\infty}^{\infty} f(x-t)\,\frac{t}{t^2+y^2}\,dt.$$

It is then an easy thing to check that $u(x,y)+iv(x,y)=F(x+iy)$, $y>0$, are the values of an analytic function on the upper half plane. An elementary argument shows that

$$\lim_{\epsilon\to 0}\left\{\int_{-\infty}^{\infty} f(x-t)\,\frac{t}{t^2+\epsilon^2}\,dt-\int_{|t|\geqq\epsilon>0} f(x-t)\,\frac{dt}{t}\right\}=0$$

almost everywhere as y tends to 0 (see Chapter 3 of [16] where this argument is

given for the conjugate function operator). Thus it suffices to show that

$$\tilde{f}(x) = \lim_{\epsilon \to 0} \frac{1}{\pi} \int_{-\infty}^{\infty} f(x - t) \frac{t}{t^2 + \epsilon^2} dt$$

exists for almost every $x \in (-\infty, \infty)$.

By considering the negative and positive parts of f separately we can reduce the problem to the case $f \leqq 0$. Thus, since $y/(t^2+y^2) > 0$ if (t, y) belongs to the upper half plane, $u \leqq 0$ and the function $G = \exp\{u+iv\} = e^F$ has absolute value $e^u \leqq 1$. By Fatou's theorem $G(z)$ has "vertical" limits a.e. on the boundary (that is, as $z = x+iy$, $y > 0$, tends to x by letting $y \to 0$). Thus, $v(z)$ must have vertical limits modulo 2π at almost all boundary points $x \in (-\infty, \infty)$. That is, the limit points of $v(z)$, as z tends vertically to x, have the form $a+2k\pi$, where k is an integer. But it is an immediate consequence of the continuity of v in the upper half plane that, if the values $v(z)$ have two distinct limit points, a and b, as z approaches x in this manner, then all points between a and b must also be limit points of this set of values. Hence, $v(z)$ can have such a limit modulo 2π only if it has a limit. This proves the first part of result (b).

In case we are dealing with an integrable function f defined on the interval $(-\pi, \pi)$ whose Fourier series is given by the expression (1.3), then the power series (1.4) does converge in the interior of the unit circle (since $|a_n|$, $|b_n|$ are dominated by $(1/2\pi) \int_{-\pi}^{\pi} |f(\theta)| d\theta$). The function F having this power series development can also be constructed by forming the Poisson integral of f which, in this case, is

$$u(re^{i\theta}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\phi) \frac{1 - r^2}{1 - 2r\cos(\theta - \phi) + r^2} d\phi$$

and the conjugate Poisson integral

$$v(re^{i\theta}) = \frac{1}{\pi} \int_{-\pi}^{\pi} f(\phi) \frac{r \sin(\theta - \phi)}{1 - 2r\cos(\theta - \phi) + r^2} d\phi.$$

We then have $F = u+iv$. The M. Riesz inequality can be obtained by any one of several complex methods. For example see Sections 2 and 3 of Chapter 7 of [**16**]. The weak type inequality (1.2) can also be obtained by making use of the properties of analytic functions (see Chapter 5 [**11**]).

It is not our intention to prove all the results stated here. Rather we would like to indicate how the theory of functions can be used to obtain them. The above argument, using Fatou's Theorem, is typical of the situation. The following more general form of Fatou's Theorem is also often used. We may express the fact that an analytic function $F(z)$ is bounded in the interior of the unit circle by asserting that

$$\mu_\infty(r) = \sup_{-\pi \leqq \theta < \pi} |F(re^{i\theta})| \leqq A < \infty,$$

for $0 \leqq r < 1$. That is, we assert that the L^∞ norms of the functions obtained by fixing r, $0 \leqq r < 1$, and whose values are $F(re^{i\theta})$, are bounded uniformly in r. We may, instead of these norms, consider other L^p norms. For example, suppose F satisfies

$$\mu_p(r) = \left(\int_{-\pi}^{\pi} |F(re^{i\theta})|^p \, d\theta\right)^{1/p} \leqq A < \infty \tag{1.5}$$

for some $p > 0$ and $0 \leqq r < 1$. It can then be shown that

$$\lim_{r\to 1} F(re^{i\theta}) = F(e^{i\theta}) \text{ exists a.e. and } \lim_{r\to 1}\left(\int_{-\pi}^{\pi} |F(re^{i\theta}) - F(e^{i\theta})|^p \, d\theta\right)^{1/p} = 0. \tag{1.6}$$

Result (a) is a consequence of this more general form of Fatou's Theorem. More precisely, it follows from the case $p = 1$.

The spaces of analytic functions satisfying condition (1.5) were first introduced by G. H. Hardy [7] and are now known as H^p spaces. Corresponding function spaces can be defined for an analytic function F in the upper half plane by saying that such a function belongs to the class H^p corresponding to this domain provided

$$\mu_p(y) = \left(\int_{-\infty}^{\infty} |F(x + iy)|^p \, dx\right)^{1/p} \leqq A < \infty,$$

for some $p > 0$ and all $y > 0$. The basic result in the theory of these spaces is that for $F \in H^p$

$$\lim_{y\to 0} F(x + iy) = F(x) \text{ a.e. and } \lim_{y\to 0} \int_{-\infty}^{\infty} |F(x + iy) - F(x)|^p \, dx = 0. \tag{1.7}$$

2. Extensions to several variables. In view of the remarks made in the last section, perhaps the most efficient way of describing how these complex methods arise in the case of harmonic analysis on Euclidean n dimensional space $\mathbf{R}^n$ is to present several possible definitions of H^p spaces of functions of several variables. The most natural extension of the classical H^p spaces of analytic functions defined in the interior of the unit circle is the space of all functions F that are holomorphic in the polydisc

$$D^n = \{z = (z_1, z_2, \cdots, z_n) \in \mathbf{C}^n : |z_1| < 1, |z_2| < 1, \cdots, |z_n| < 1\}$$

and satisfy

$$\int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} |F(r_1e^{i\theta_1}, \cdots, r_ne^{i\theta_n})|^p d\theta_1 \cdots d\theta_n \leqq A < \infty$$

for $0 \leqq r_1, r_2, \cdots, r_n < 1$. The theory of these spaces was first developed by A. Zygmund [15]. He showed that (1.6) has the following extension

$$\lim_{r_1, r_2, \cdots, r_n \to 1} F(r_1e^{i\theta_1}, r_2e^{i\theta_2}, \cdots, r_ne^{i\theta_n}) = F(e^{i\theta_1}, e^{i\theta_2}, \cdots, e^{i\theta_n}),$$

exists for almost every $(\theta_1, \theta_2, \cdots, \theta_n) \in [-\pi, \pi] \times \cdots \times [-\pi, \pi]$ and

$$\lim_{r_1, r_2, \cdots, r_n \to 1} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} | F(r_1 e^{i\theta_1}, \cdots, r_n e^{i\theta_n}) - F(e^{i\theta_1}, \cdots, e^{i\theta_n}) |^p d\theta_1 \cdots d\theta_n = 0.$$

We note that the entire boundary of the polydisc is not involved. Rather, only the product of n perimeters of unit circles is being used as the domain of definition of the boundary values.

The analogous extension of the spaces defined at the very end of Section 1 is the class of all holomorphic functions F defined in the product of half planes $\{z = (x_1 + iy_1, \cdots, x_n + iy_n) : y_1 > 0, \cdots, y_n > 0\}$ satisfying

$$\int_{-\infty}^{\infty} | F(x_1 + iy_1, \cdots, x_n + iy_n) |^p dx_1 \cdots dx_n \leqq A < \infty$$

for all $y_1, y_2, \cdots, y_n > 0$. Again, it can be shown that for such a function F we have (see [16])

$$\lim_{y_1, y_2, \cdots, y_n \to 0} F(x_1 + iy_1, x_2 + iy_2, \cdots, x_n + iy_n) = F(x_1, x_2, \cdots, x_n)$$

exists for a.e. $x = (x_1, \cdots, x_n) \in \mathbf{R}^n$ and

$$\lim_{y_1, y_2, \cdots, y_n \to 0} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} | F(x_1 + iy_1, \cdots, x_n + iy_n) - F(x_1, \cdots, x_n) |^p dx_1 \cdots dx_n = 0.$$

The more complicated structure of $\mathbf{R}^n$, when $n > 2$, however, allows us to consider much more general function spaces that directly extend the classical H^p spaces. For example, one can obtain such a function space associated with every tube domain in the complex n-dimensional spaces $\mathbf{C}^n$. These domains are defined in the following way: Let B be an open convex subset of $\mathbf{R}^n$ then the tube T_B having base B is the domain

$$T_B = \{x + iy : x \in \mathbf{R}^n, y \in B\}.$$

Then a function F is said to belong to the space $H^p(T_B)$, $p > 0$, provided it is holomorphic in T_B and satisfies

$$\int_{\mathbf{R}^n} | F(x + iy) |^p dx \leqq A < \infty \tag{2.1}$$

for all $y \in B$. When the dimension n is 1 and $B = (0, \infty)$, $H^p(T_B)$ is the space defined at the end of the last section. Since we are assuming that B is convex, the only other possible types of tube domains are those with base a finite interval or the entire real line. (If B is merely open and connected, but not neces-

sarily convex, and F is a holomorphic function in T_B satisfying (2.1) then F has a unique holomorphic extension to T_{B^c}, where B^c is the convex hull of B, which satisfies (2.1) for all $y \in B^c$ (see [**11**]).) In the latter case it follows from condition (2.1) that $F \in H^p(T_B)$ implies $F \equiv 0$. In the former case we obtain an interesting function space having many of the features of the better known classical space associated with the upper half-plane.

One would expect, in view of result (1.7), that, in general, for $F \in H^p(T_B)$ and y_0 a boundary point of B, $F(x+iy)$ tends to a limit function $F(x+iy_0)$ as $y \in B$ tends to y_0 (either almost everywhere or in the mean). Unfortunately, when $n>1$ this is not true even when $p=2$ and we consider only limits in the mean. (These are the simplest H^p spaces one can study since many of their properties can be obtained by making use of the behavior of the Fourier transform as an operator in $L^2(\mathbf{R}^n)$. In particular, many of these properties are consequences of the Plancherel theorem. Bochner ([**1**] and [**2**]) was the first to introduce the spaces $H^2(T_B)$. The more general $H^p(T_B)$ spaces, $0<p$, were first introduced in [**10**].) In order to assure the existence of such limit functions we must restrict the approach of y to the boundary point y_0 to be "nontangential" (see [**10**] for a precise definition of this notion). Despite such difficulties the H^2 theory is sufficiently rich to enable one to obtain several applications. For example, it can be used to derive a characterization of the Fourier transforms of L^2 functions vanishing outside certain compact and convex subsets of $\mathbf{R}^n$ (this is an extension of the Paley-Wiener theorem; see the third chapter of [**10**] for details).

By restricting the base of the tube domain we can obtain other direct extensions of results in the classical one-dimensional theory. For example, let Γ be an *open cone* with vertex at the origin; that is, Γ is an open set in $\mathbf{R}^n$ having the property that whenever $x, y \in \Gamma$ and $a, b>0$ then $ax+by \in \Gamma$. We say that Γ is *regular* if it does not contain any entire straight lines. Associated with a tube domain T_Γ, with Γ a regular open cone, one can find a Poisson kernel (see [**10**]) which makes it possible to obtain an H^p space theory for all $p>0$. One of the applications of this theory is the following extension of result (a):

If μ is a finite Borel measure on $\mathbf{R}^n$ whose Fourier-Stieltjes transform has support in a regular open cone then μ is absolutely continuous.

The proof of this result is very similar to the classical one: The hypotheses imply that the Poisson integral of the measure gives us a function belonging to $H^1(T_{\Gamma^*})$, where Γ^* is a cone associated with Γ (called the "dual" to Γ). Once this is done, the rest of the proof is essentially the same as that of the 1-dimensional result. (This application was brought to the author's attention in a conversation with Professor R. R. Coifman.)

The theory of H^p spaces becomes much richer if we consider certain special cones. For example, if Γ is the circular cone $\{y=(y_1, y_2, \cdots, y_n) \in \mathbf{R}^n : y_1^2 - y_2^2 - \cdots - y_n^2 > 0 \text{ and } y_1 > 0\}$ the Poisson kernel has the following particularly simple form for $(x, y) \in T_\Gamma$:

$$P_y(x) = \frac{c(y, y)^{n/2}}{([(x, x) - (y, y)]^2 + 4(x, y)^2)^{n/2}},$$

where $(u, v) = u_1v_1 - u_2v_2 - \cdots - u_nv_n$ and c is an appropriate constant. Similar formulae can be obtained for the Poisson kernel associated with other cones that are "classical domains." These expressions make it possible to obtain certain estimates for Poisson integrals useful in extending one-dimensional results. For details see Chapter 3 of [11].

One can extend the H^p theory associated with the unit disc to n dimensions by a method analogous to the one just described. This involves the use of Reinhardt regions instead of tube domains. The problems arising from this development are similar to those encountered in the H^p space theory associated with tubes.

There exist other ways of considering n-dimensional extensions of complex methods. We shall finish this article by describing certain H^p spaces for which the basic novelty in the generalization is not in the domain used, but in what is meant by an analytic function of n variables. If $F = u + iv$ is analytic in some region of the plane $\mathbf{R}^2$, we know that the Cauchy-Riemann equations $v_x + u_y = 0$ and $v_y - u_x = 0$ must be satisfied in this region. Conversely, if two differentiable functions u and v satisfy these equations in a simply connected region $U \subset \mathbf{R}^2$, then $F = u + iv$ is analytic in U. The second equation asserts that (v, u) is the gradient of a function h, while the first assures us that h is harmonic. Consequently, $F = u + iv$ is analytic if and only if (v, u) is the gradient of a harmonic function h in every simply connected subregion of the domain of F.

In view of these observations it is reasonable to consider the gradient of a harmonic function defined in a domain $U \subset \mathbf{R}^n$ to be a type of "generalized analytic function." If $F = (u_1, u_2, \cdots, u_n)$ is such a gradient then, clearly, the components $u_1, u_2, \cdots, u_n$ satisfy the system of partial differential equations

$$\frac{\partial u_1}{\partial x_1} + \frac{\partial u_2}{\partial x_2} + \cdots + \frac{\partial u_n}{\partial x_n} = 0, \qquad \frac{\partial u_j}{\partial x_k} - \frac{\partial u_k}{\partial x_j} = 0, \tag{2.2}$$

for $j, k = 1, 2, \cdots, n$. We call a solution $F = (u_1, u_2, \cdots, u_n)$ of (2.2) in a connected (not necessarily simply connected) region $U \subset \mathbf{R}^n$ an *M. Riesz system of conjugate harmonic functions*. (A simple calculation shows that any differentiable solution of (2.2) consists of functions that are harmonic.) Such a vector valued function F is an example of the generalization of the notion of analytic function we mentioned in the previous paragraph.

We can define the H^p space ($p > 0$) associated with the upper half space $\mathbf{R}^{n+1}_+ = \{(x, y) = (x_1, x_2, \cdots, x_n, y) \in \mathbf{R}^{n+1} : x \in \mathbf{R}^n, y > 0\}$ to consist of all those *M. Riesz systems of conjugate harmonic functions* $F = (u_1, u_2, \cdots, u_{n+1})$ defined in $\mathbf{R}^{n+1}_+$ which satisfy

$$\left(\int_{\mathbf{R}^n} |F(x, y)|^p dx \right)^{1/p} \leqq A < \infty \tag{2.3}$$

for all $y>0$, where $|F| = (u_1^2+u_2^2+ \cdots +u_{n+1}^2)^{1/2}$. We see that these spaces also furnish us with a simple direct generalization of those introduced at the end of the last section. Unfortunately, we do not obtain a complete extension of (1.7). The best result that has been established on the existence of boundary values for $F\in H^p(\mathbf{R}_+^{n+1})$ (see [12]) is that for $p \geqq (n-1)/n$

$$\lim_{y\to 0} F(x, y) = F(x) \tag{2.4}$$

exists for a.e. $x\in \mathbf{R}^n$ and

$$\lim_{y\to 0}\left(\int_{\mathbf{R}^n} | F(x, y) - F(x) |^p dx\right)^{1/p} = 0 \tag{2.5}$$

when $p>(n-1)/n$.

The significant feature of this result is that we do obtain boundary values for indices $p\leqq 1$. (If $\infty > p > 1$ and u is a harmonic function on $\mathbf{R}_+^{n+1}$ satisfying $(\int_{\mathbf{R}^n} |u(x, y)|^p dx)^{1/p} \leqq A < \infty$ for all $y>0$ then the boundary values $\lim_{y\to 0} u(x, y) = u(x)$ exist for almost every $x\in\mathbf{R}^n$ and $\lim_{y\to 0}(\int_{\mathbf{R}^n}|u(x, y)-u(x)|^p dx)^{1/p}=0$. When $p=1$, the last equality fails in general but the almost everywhere defined "vertical" boundary values do exist. When $p<1$ even these boundary values may fail to exist (see [11]). By considering a system of conjugate harmonic functions the results we have just announced tell us that we obtain boundary values for p below 1, the index which is so significant in determining the boundary behavior of a single harmonic function.) For example, this fact allows to establish yet another extension of the theorem of F. and M. Riesz (see [12] for a formulation of this extension). A study of the proof of (2.4) and (2.5) shows that the basic reason for being limited to values $p \geqq (n-1)/n$ is that a system $F=(u_1, u_2, \cdots, u_{n+1})$ satisfying equations (2.2) has the property that $|F|^p$ is *subharmonic* for $p\geqq (n-1)/n$. In our case this is equivalent to saying that $\Delta|F|^p \geqq 0$ whenever $F(x, y)\neq 0$, where Δ is the Laplace operator $(\partial^2/\partial x_1^2) + \cdots + (\partial^2/\partial x_n^2)+(\partial^2/\partial y^2)$. It turns out that the range of indices p for which $|F|^p$ is subharmonic is larger for other systems of conjugate harmonic functions. For each such system we have corresponding H^p classes, defined by (2.3), and boundary value results (2.4) and (2.5). We shall end this exposition by presenting some of these systems.

We introduced the M. Riesz system as a gradient $\nabla h = F = (u_1, u_2, \cdots, u_n)$ of a harmonic function h. By considering the gradient $\nabla u_j = (u_{j1}, u_{j2}, \cdots, u_{jn})$ of each of the components u_j, $j=1, 2, \cdots, n$, we obtain an n^2-tuple $F^{(2)} = (u_{11}, u_{12}, \cdots, u_{nn})$ which we shall call the *second gradient of h*. We can perform this operation again on each of the components of $F^{(2)}$ and obtain the *third gradient* of h. Continuing in this way we obtain the kth *gradient* $F^{(k)}$ *of h*. Calderón and Zygmund [4] have shown that $|F^{(k)}|^p$ is subharmonic for $p\geqq (n-2)/(k+n-2)$. Thus, by considering H^p spaces of systems of harmonic functions on $\mathbf{R}_+^{n+1}$ that are kth gradients we obtain the boundary value results (2.4) and (2.5) for indices $p\geqq (n-1)/(k+n-1)$ (observe that the dimension of

the domain is $n+1$). That is, we obtain an H^p space theory for indices p arbitrarily close to 0.

These higher gradients satisfy certain partial differential equations which can be included in a class of equations closely connected with the irreducible representations of the rotation group $SO(n)$ and its covering group (see [13]). Even more general systems of partial differential equations can be written in the form

$$\sum_{j=1}^{n} A_j \frac{\partial F}{\partial x_j} = 0, \tag{2.6}$$

where A_j is an $m\times k$ constant matrix, $F=(u_1, u_2, \cdots, u_k)$ is a k-tuple, and $\partial F/\partial x_j=(\partial u_1/\partial x_j, \partial u_2/\partial x_j, \cdots, \partial u_k/\partial x_j)$. If each solution of (2.6) must be harmonic, these partial differential equations are known as *generalized Cauchy-Riemann* systems. It can be shown (see [5]) that for each such system there exists an index $p_0<1$ such that, if F is a solution then $|F|^p$ is subharmonic for $p\geqq p_0$. It would be interesting to know which of these generalized Cauchy-Riemann systems are the ones known to be associated with the irreducible representations mentioned above.

References

1. S. Bochner, Group invariance of Cauchy's formula in several variables, Ann. of Math., 45 (1944) 686–707.

2. ———, Classes of holomorphic functions of several complex variables in circular domains, Proc. Nat. Acad. Sci. U.S.A., 46 (1960) 721–723.

3. A. P. Calderón, Commutators of singular integral operators, Proc. Nat. Acad. Sci. U.S.A., 53 (1965) 1092–1099.

4. ———, and A. Zygmund, On higher gradients of harmonic functions, Studia Math., 24 (1964) 211–226.

5. R. R. Coifman and G. Weiss, On subharmonicity inequalities involving solutions of generalized Cauchy-Riemann equations, to appear in Studia Math., 36 (1970).

6. P. Fatou, Séries trigonométriques et séries de Taylor, Acta Math., 30 (1906) 335–400.

7. G. H. Hardy, The mean value of the modulus of an analytic function, Proc. London Math. Soc., 14 (1914) 269–77.

8. H. Helson, Proof of a conjecture of Steinhaus, Proc. Nat. Acad. Sci. U.S.A., 40 (1954) 205–6.

9. F. Riesz and M. Riesz, Über Randwerte einer analytischen Funktion, Quatrième Congrès des mathématiciens scandinaves, Stockholm, 1916, 27–44.

10. E. M. Stein, M. Weiss and G. Weiss, H^p classes of holomorphic functions in tube domains, Proc. Nat. Acad. Sci. U.S.A., 52 (1964) 1035–1039.

11. ———, and G. Weiss, Introduction to Fourier analysis on n-dimensional Euclidean space, to be published by the Princeton Univ. press in 1970.

12. ——— and ———, On the theory of harmonic functions of several variables, Acta Math., 103 (1960) 26–62.

13. ——— and ———, Generalization of the Cauchy-Riemann equations and representations of the rotation group, Amer. J. Math., 90 (1968) 163–196.

14. G. Weiss, Harmonic Analysis, MAA Studies in Mathematics, Vol. III, (1965) 124–178.

15. A. Zygmund, On the boundary values of functions of several complex variables, Fund. Math., 36 (1949) 207–235.

16. ———, Trigonometric Series, Vols. I and II, 2nd ed., Cambridge University Press, 1968.

CRITICAL POINTS AND CURVATURE FOR EMBEDDED POLYHEDRAL SURFACES

T. F. BANCHOFF, Brown University

The Gauss-Bonnet theorem for a surface M^2 in Euclidean 3-space E^3 and the Critical Point Theorem for height functions on an embedded surface are two of the earliest and most important theorems of "geometry in the large." Both theorems relate geometric properties of the embedded surface (the total curvature of the surface or the sum of a set of geometrically defined indices of singularity) to a topological property of the surface, the Euler-Poincaré characteristic $\chi(M^2)$. Both theorems are very geometric in character despite the fact that the standard definitions of total curvature and index of singularity appear to involve the use of differential calculus and the hypothesis that the surfaces are smooth. In fact both theorems have analogues for polyhedral surfaces embedded in E^3, and the proofs in the polyhedral case are entirely elementary.

Some of the most interesting results in global geometry have exploited the connection between total curvature and critical point theory, as in the work of

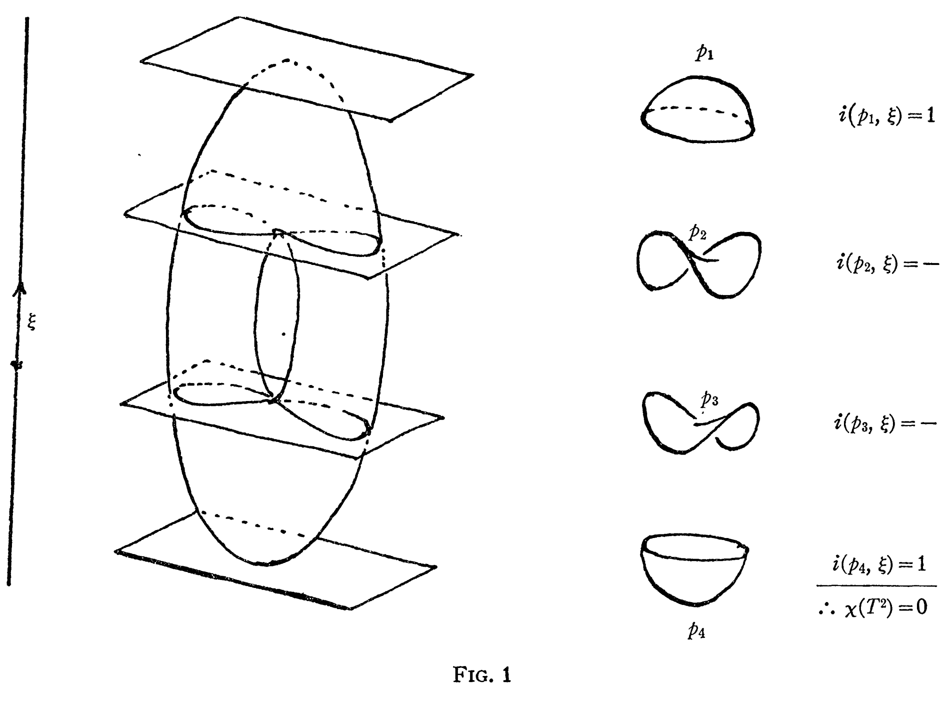


FIG. 1

Thomas Banchoff wrote his Berkeley thesis under Chern in 1964. He was a Peirce Instructor at Harvard, a Fulbright Research Associate at the University of Amsterdam, and presently is an Assistant Professor at Brown. His research is in global differential geometry. *Editor.*

Kuiper [4]. In this paper we shall follow this same procedure to prove the critical point theorem and use it to prove the Gauss-Bonnet theorem. In the polyhedral case, a new feature is an interpretation of the *Theorema Egregium* of Gauss which relates the extrinsic curvature to the intrinsic curvature on a surface.

All of the theorems in this paper have appeared in a generalized (and technical) form in the author's paper [2].

1. The Critical Point Theorem. Consider a closed smooth surface M^2 embedded in E^3 and consider a linear function ξ on E^3 given by projecting all of E^3 to the line determined by a unit vector ξ. A point p of M^2 is said to be a *critical point* for ξ if the tangent plane to M^2 at p is perpendicular to ξ; all other points of M^2 are called *ordinary points* for ξ. In the "standard" example of a height function on a torus of revolution held vertically there are just four critical points, a maximum, a minimum, and two (nondegenerate) saddle points.

The *Critical Point Theorem* for height functions states that if ξ has a finite number of critical points on M^2 and all are of the three types described above, then (number of local maxima) + (number of local minima) − (number of saddle points) $=\chi(M^2)$, where $\chi(M^2)$ is the Euler-Poincaré characteristic of M^2.

We express this theorem more succinctly by indexing each critical point by $i(p,\xi)=1$ if p is a local maximum or minimum and $i(p,\xi)=-1$ if p is a (nondegenerate) saddle point. The theorem then states

$$\sum_{p \text{ critical for } \xi} i(p,\xi) = \chi(M^2).$$

In classical critical point theory (= Morse Theory) the index is given by considering the sign of the determinant of the matrix of second derivatives, as in [5], but since we are interested in developing the polyhedral analogue of the theorem, we proceed to give a more geometric presentation of this indexing procedure.

If a point q is ordinary for the height function ξ, then the tangent plane at q is not horizontal. Thus the tangent plane divides a "small disc neighborhood" U of q on M^2 into exactly two pieces and it meets a "small circle" about q in precisely two points. This distinguishes an ordinary point from a local maximum or minimum (where a "small circle" about the critical point will not meet the tangent plane at all) and from a nondegenerate saddle point p (where the plane at p perpendicular to q meets a "small circle" about p on M^2 in four distinct points).

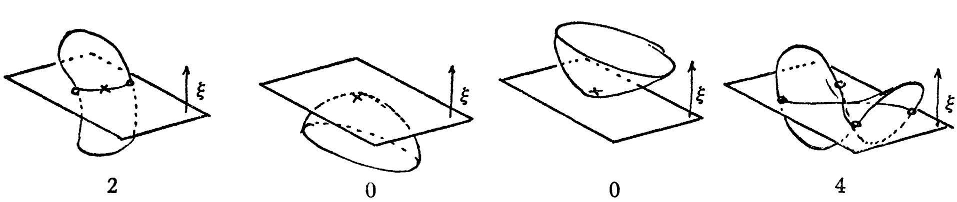


FIG. 2

We may then give an arithmetic definition of the *index* as follows:

$i(p, \xi) = 1 - \frac{1}{2}$ (number of points in which the plane through p perpendicular to ξ meets a "small circle" about p on M^2).

This definition agrees with the previous indexing procedure and has the additional property that $i(q, \xi) = 0$ if q is not a critical point. In the smooth case, however, the definition is somewhat unsatisfactory due to the difficulty of defining precisely the notion of a "small circle." In the polyhedral case, on the other hand, this is exactly the sort of definition which we want.

Consider a polyhedral surface M^2 in E^3 which is expressed as a union of V vertices, E edges, and T triangular faces. The *Euler characteristic* of M^2 is defined to be

$$\chi(M^2) = V - E + T.$$

A height function ξ on E^3 is said to be *general for the polyhedral surface* M^2 if $\xi(v) \neq \xi(w)$ whenever v and w are distinct vertices of M^2. If ξ is general for M^2,

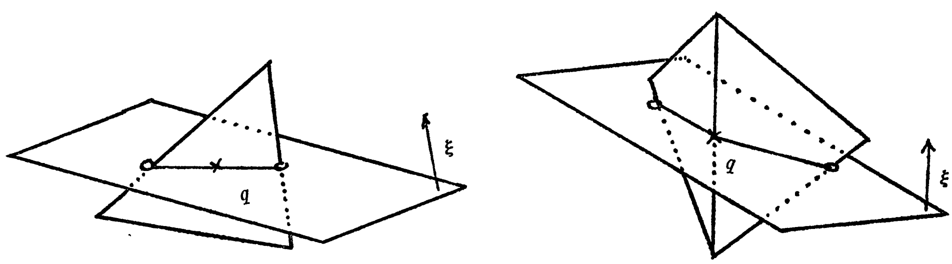


FIG. 3

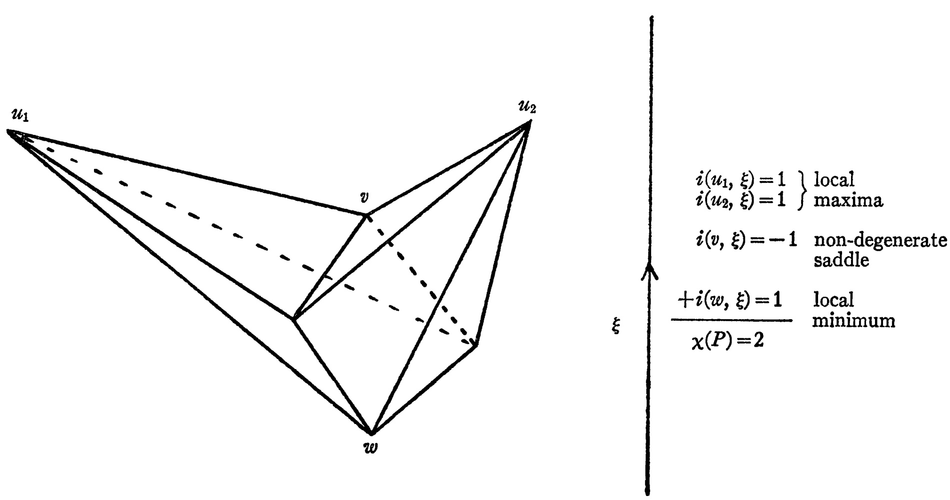
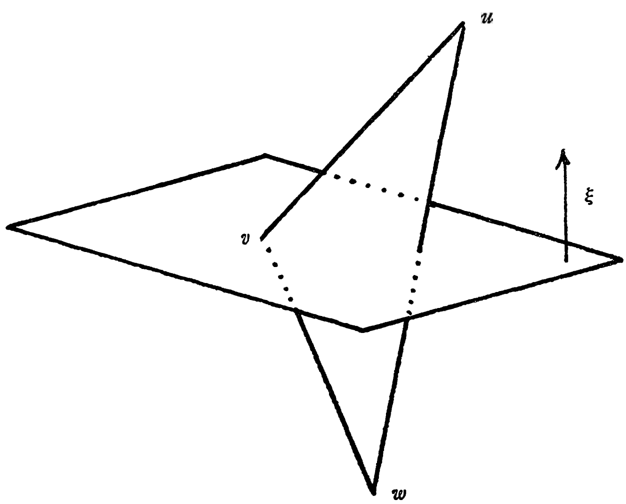


FIG. 4

then the point q is said to be *ordinary* for ξ if the plane through p perpendicular to ξ cuts the disc neighborhood Star(q) into two pieces, where Star(q) is the union of all vertices, edges, and faces which include q. (When we say M^2 is a polyhedral surface, we mean that for each point q, Star(q) is the image of an open disc in the plane under a one-to-one continuous map.) With this definition, any point q in the interior of a face or an edge has to be ordinary since no face or edge can be perpendicular to the vector ξ if ξ is general for M^2.

For vertices, however, there are critical points corresponding to all the types presented for smooth functions, as, for example, in the indicated polyhedron. We may then use the indexing procedure developed for smooth surfaces, where instead of a "small circle" we use the embedded polygon which is the boundary of the star of the vertex v. The number of times the plane through v perpendicular to ξ meets this polygon is then equal to the number of triangles $\triangle$ in Star(v) such that one of the vertices of $\triangle$ lies above the plane and the other lies below.



v is the middle vertex of $\triangle$ for ξ.

FIG. 5

In such a case we say that v is the middle vertex of $\triangle$ for ξ, and we may write the index as follows:

$$i(v, \xi) = 1 - \tfrac{1}{2} \text{ (number of } \triangle \text{ with } v \text{ middle for } \xi).$$

Again this definition corresponds to the definitions given in the smooth case and it gives index 0 for an ordinary point.

The critical point theorem then states:

THEOREM 1. *If ξ is general for M^2, then*

$$\sum_{v \in M^2} i(v, \xi) = \chi(M^2).$$

We require a lemma on polyhedral surfaces:

LEMMA. *For a polyhedral surface*, $3T=2E$.

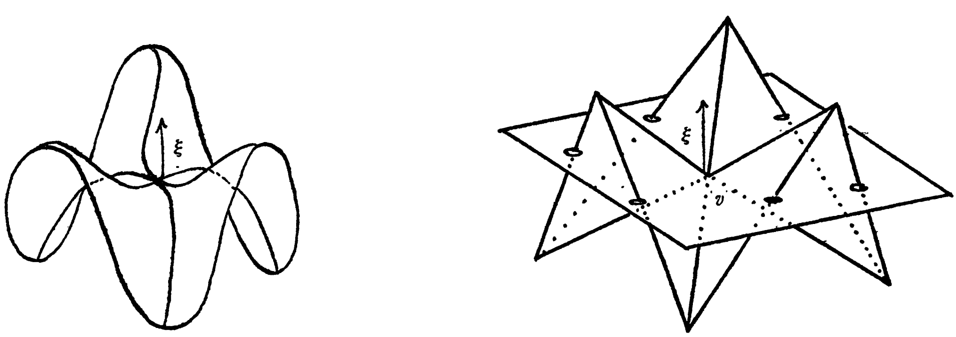
Proof of Lemma. Since an edge in a polyhedral surface has precisely two triangles in its star,

$$3T = \text{number of pairs } (\triangle, \text{ edge of } \triangle) = 2E.$$

Proof of Theorem. If ξ is general for M^2,

$$\begin{aligned}\sum_{v\in M} i(v,\xi) &= \sum_{v\in M} (1 - \tfrac{1}{2}(\text{number of } \triangle \text{ with } v \text{ middle for } \xi)),\\ &= V - \tfrac{1}{2}\sum_{v\in M}(\text{number of } \triangle \text{ with } v \text{ middle for } \xi),\\ &= V - \tfrac{1}{2}T \text{ (since each } \triangle \text{ has exactly one middle vertex for } \xi),\\ &= V - \tfrac{1}{2}(2E-2T) \text{ (since } T = 2E - 2T \text{ by the lemma)},\\ &= V - E + T.\end{aligned}$$

REMARK. For a smooth surface M^2 embedded in E^3 it is a classical result that for almost every unit vector ξ on S^2 (i.e., except for a set of measure zero on S^2), the height function ξ has only finitely many critical points, and furthermore, for almost all ξ, this height function has as critical points only local maxima and minima, and nondegenerate saddle points. In the polyhedral case it is immediate that almost all ξ are general for M^2 (since the nongeneral ξ lie in the finite union of great circles $\{\xi\in S^2 \mid \xi(v)=\xi(w)\}$, one for each pair of distinct vertices v, w). The stronger result, however, is not correct in the polyhedral case. Consider an isolated critical point of a smooth surface which is degenerate—the "monkey saddle" (so called in Hilbert and Cohn-Vossen [3], p. 191, since a monkey riding a bicycle would need three depressions in his saddle, one for



$i(v, \xi) = -2$

FIG. 6

each leg and one for his tail). Although on the smooth monkey saddle, height functions η near ξ on S^2 have only nondegenerate critical points, any η near ξ

has exactly six triangles with v middle, so for each such η, $i(v, \eta) = -2$.

However, in the proof of Theorem 1, we never required that the indices be only 0, 1 or -1. The proof goes through without change for any polyhedral surface M and ξ general for M, regardless of the complexity of the stars of the vertices of M.

2. Total Curvature and the Gauss-Bonnet Theorem. The total curvature or Gaussian curvature of a neighborhood U on a smooth surface M^2 in E^3 has several definitions which appear in texts in differential geometry. The definition which most easily leads to an extrinsic curvature theory for polyhedra is the one originally given by Gauss. We sketch his procedure in the smooth case and then develop the analogous theory for arbitrary embedded polyhedral surfaces in E^3.

Consider a "small" neighborhood U_1 on a convex surface M^2 in E^3. The Gauss map $g: U_1 \to S^2$ is defined by setting $g(p) =$ outward unit normal vector to M^2 at p. If the mapping g restricted to U_1 is one-to-one, and g is orientation-preserving (outward normals at corresponding points correspond) then U_1 is said to be *strictly convex*. The *total curvature* $\tilde{K}(U_1)$ of U_1 is then defined to be the area of the spherical image $g(U_1)$ on S^2.

If U_2 is a region of a nonconvex surface M^2 on which g is one-to-one and orientation-reversing, then $\tilde{K}(U_2)$ is defined to be the negative of the area of $g(U_2)$.

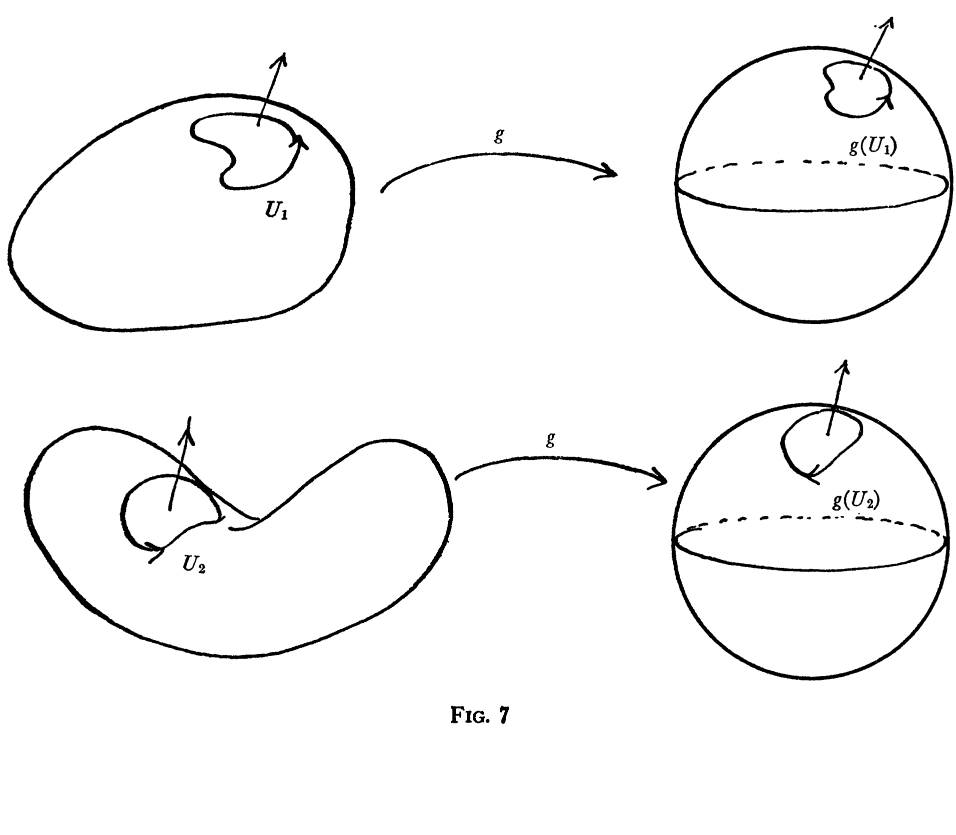


FIG. 7

We may use the index $i(p, \xi)$ to describe the total curvature of U_1 in these cases. Let $d\omega$ denote the area integrand for S^2, so

$$\int_{S^2} d\omega = 4\pi.$$

Then the area of $g(U_1)$ is the integral over S^2 of the characteristic function: $F_{g(U_1)}(\xi)=1$ if ξ is in $g(U_1)$ and 0 otherwise. But if ξ is in $g(U_1)$, then the height function ξ has a critical point at a point p on U_1. In fact, $i(p, \xi)\neq 0$ for a point p of U_1 if and only if ξ is in $g(U_1)$ or $-\xi$ is in $g(U_1)$. We may assume that U_1 is small enough that $g(U_1)$ contains no pair of antipodal points. The total curvature may then be described as follows:

$$\tilde{K}(U_1) = \int_{S^2} F_{g(U_1)}(\xi)d\omega = \frac{1}{2}\int_{S^2} \sum_{p\in U_1} i(p, \xi)d\omega.$$

But the right-hand expression also serves to define $\tilde{K}(U_2)$, since this expression gives the negative of the area of $g(U_2)$.

The following two paragraphs explain the procedure of defining the total curvature in the smooth case, and this serves to motivate the definition for the polyhedral analogue. In the polyhedral case, however, the technical difficulties concerning the convergence of the integrand do not occur.

In order to define the total curvature of a neighborhood U on which g is not one-to-one, we begin by expressing U as a countable disjoint union of sets U_i on which g is one-to-one together with a set $V=U-\bigcup_{i=1}^{\infty} U_i$ such that $g(V)$ has measure zero on S^2. Then we set $\tilde{K}(U)=\sum_{i=1}^{\infty} \tilde{K}(U_i)$ if this sum converges, and we obtain a totally additive set function on M^2. This definition then coincides with

$$\tilde{K}(U) = \frac{1}{2}\int_{S^2} \sum_{p\in U} i(p, \xi)d\omega,$$

where the integrand is well defined almost everywhere since almost every ξ has only finitely many critical points.

REMARK. In the case that M^2 is sufficiently smooth, the set function $\tilde{K}(U)$ is absolutely continuous with respect to the area measure $A(U)$ on M^2, and we may define the point function $K(p)$ as the limit of $\tilde{K}(U)/A(U)$ for any collection of neighborhoods U_i of p with the limit of the diameters of U_i going to zero. This function is called the *Gaussian curvature* at p, and by integrating this function with respect to the area measure, we obtain

$$\tilde{K}(U) = \int_U K(p)dA.$$

The classical Gauss-Bonnet theorem for embedded closed surfaces states that

$$\tilde{K}(M^2) = \int_{M^2} K(p)dA = 2\pi\chi(M^2).$$

In the case of a polyhedral manifold M^2 in E^3, we may use the same definition as that developed for smooth surfaces. For any open set U of M^2, we set

$$K(U) = \frac{1}{2}\int_{S^2} \sum_{v\in U} i(v, \xi)d\omega.$$

When M^2 is a convex polyhedron, if U is a neighborhood containing only one vertex v, then $\tilde{K}(U)$ gives the measure of the exterior angle at v, i.e., the set of normals to support planes to M^2 at v, and this approach has been used in the classical theory of convex polyhedra, for example in [1].

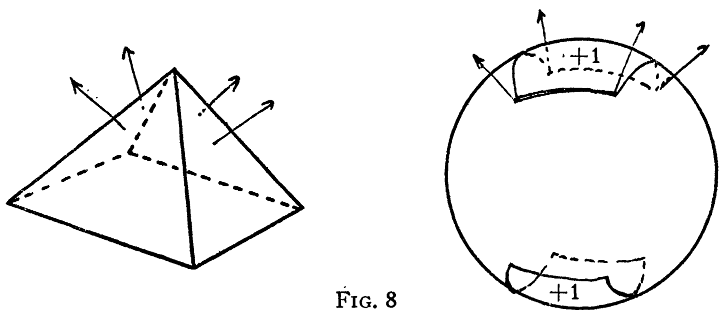


FIG. 8

The expression above also yields a definition of curvature for nonconvex vertices of saddle type as well as for the monkey saddles.

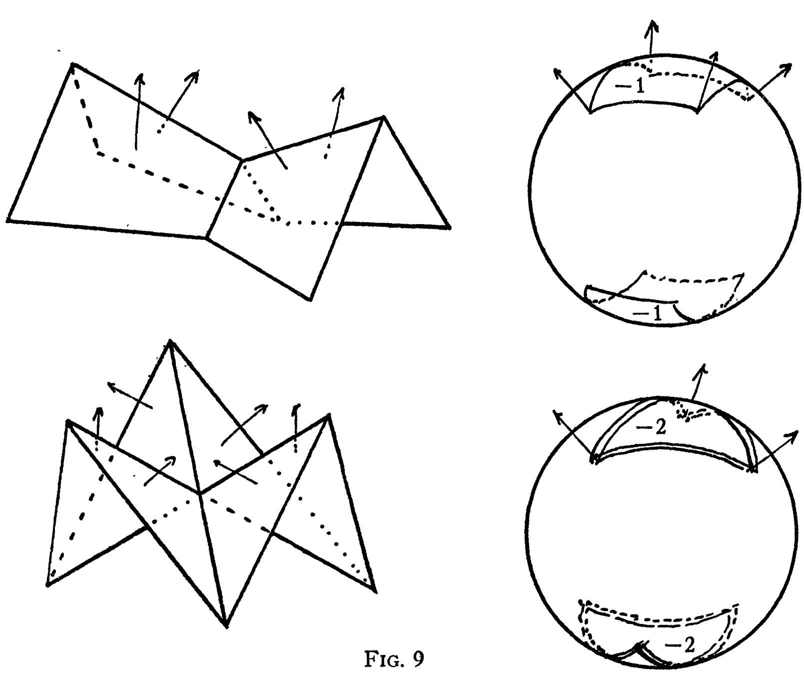


FIG. 9

The Gauss-Bonnet theorem for embedded surfaces, smooth or polyhedral, follows immediately:

THEOREM 2. $\tilde{K}(M^2)=2\pi\chi(M^2)$.

Proof.

$$\tilde{K}(M^2) = \frac{1}{2}\int_{S^2}\sum_{p\in M^2} i(p,\xi)d\omega = \frac{1}{2}\int_{S^2}\chi(M^2)d\omega = \frac{1}{2}\chi(M^2)\int_{S^2}d\omega = 2\pi\chi(M^2).$$

3. The Theorema Egregium. Total curvature was defined for a set U on a smooth surface M in a way which involved the extrinsic properties of the surface, that is, the way M is situated in E^3. Gauss, however, proved that in fact $\tilde{K}(U)$ depends only on intrinsic properties of U, i.e., on properties that are determined by measurements made along the surface not taking into account the way the surface is situated in space, and Gauss called this result his *Theorema Egregium*.

We shall prove the analogous theorem for embedded polyhedral surfaces. Observe first of all that for an open set U of M,

$$\tilde{K}(U) = \frac{1}{2}\int_{S^2}\sum_{v\in U} i(v,\xi)d\omega = \sum_{v\in U}\frac{1}{2}\int_{S^2} i(v,\xi)d\omega$$

so we may set

$$\tilde{K}(U) = \sum_{v\in U}\tilde{K}(v), \quad \text{where} \quad \tilde{K}(v) = \frac{1}{2}\int_{S^2} i(v,\xi)d\omega.$$

THEOREM 3. *$\tilde{K}(v)$ is intrinsic, in fact, $\tilde{K}(v)=2\pi-$(sum of interior angles at v of triangles containing v).*

Proof. Let $m(v,\triangle,\xi)$ be the function on E^3 defined by $m(v,\triangle,\xi)=1$ if v is the middle vertex of $\triangle$ for ξ and $m(v,\triangle,\xi)=0$ otherwise. Then $i(v,\xi)=1-\frac{1}{2}\sum_{\triangle\in M} m(v,\triangle,\xi)$ and

$$\begin{aligned}\tilde{K}(v) &= \frac{1}{2}\int_{S^2} i(v,\xi)d\omega = \frac{1}{2}\int_{S^2}\left(1-\frac{1}{2}\sum_{\triangle\in M} m(v,\triangle,\xi)\right)d\omega\\ &= \frac{1}{2}\int_{S^2}d\omega - \frac{1}{4}\int_{S^2}\sum_{\triangle\in M} m(v,\triangle,\xi)d\omega = 2\pi - \sum_{\triangle\in M}\frac{1}{4}\int_{S^2} m(v,\triangle,\xi)d\omega.\end{aligned}$$

The proof is then complete once we establish the following lemma:

LEMMA. $\int_{S^2} m(v,\triangle,\xi)d\omega = 4$ *(interior angle of $\triangle$ at v).*

Proof. First of all, observe that for vectors $\mathbf{n}$ at v in the plane E^2 containing $\triangle$, $m(v,\triangle,\eta)=1$ if and only if $\mathbf{n}$ lies in the region between the lines perpendicular to the edges $\boldsymbol{u}-\boldsymbol{v}$ and $\boldsymbol{w}-\boldsymbol{v}$ and the angle which determines this region is equal to the interior angle of $\triangle$ at $\boldsymbol{v}$. Any vector $\boldsymbol{\xi}$ of S^2 may be written uni-

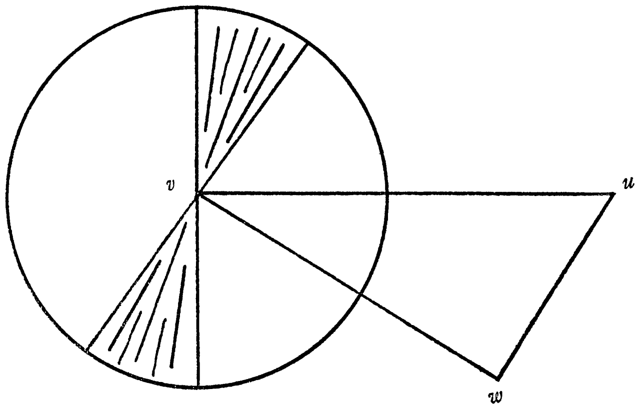


FIG. 10

quely as $\xi=\mathbf{n}+\zeta$ where $\mathbf{n}$ lies in the plane E^2 containing $\triangle$ and ζ is perpendicular to E^2. Then $m(v, \triangle, \xi)=1$ if and only if $\xi\cdot(u-v)>0>\xi\cdot(v-w)$ or $\xi\cdot(u-v)<0<\xi\cdot(v-w)$ so $m(v, \triangle, \xi)=1$ if and only if $m(v, \triangle, \eta)=1$. Thus the set of ξ on S^2 centered at v for which $m(v, \triangle, \xi)=1$ forms a double lune with axis perpendicular to E^2 and with each angle equal to the interior angle of $\triangle$ at v. But the area of a lune is twice the angle of the lune, so

$$\int_{S^2} m(v, \triangle, \xi)d\omega = 4 \text{ (interior angle of } \triangle \text{ at } v).$$

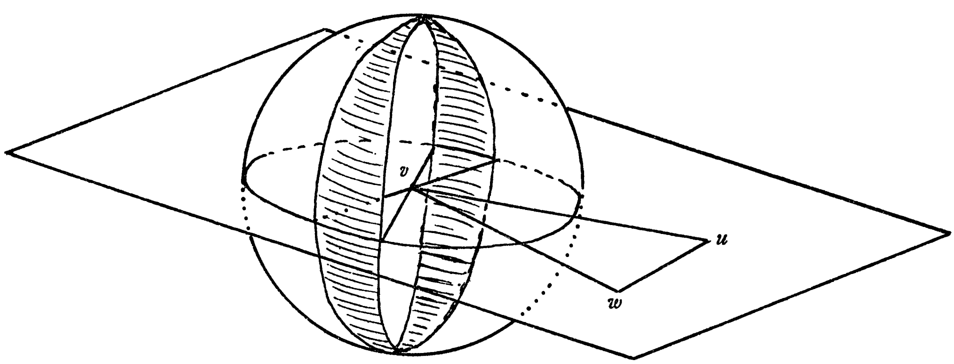


FIG. 11

This completes the proof of the *Theorema Egregium* for polyhedra, no matter how complicated the vertex stars are. Compare Hilbert and Cohn-Vossen ([**3**], p. 195) for a similar argument for vertices corresponding to nondegenerate saddle points.

REMARK. The theorems of section 2 may be considered as a generalization of the approach of G. Pólya for embedded convex polyhedral discs [**6**].

References

1. A. D. Alexandroff, Konvexe Polyeder, Akademie Verlag, Berlin, 1958 (Russian original: Moscow, 1950).
2. T. F. Banchoff, Critical Points and Curvature for Embedded Polyhedra, J. of Differential Geometry, number 3, 1 (1967) 257–268.
3. D. Hilbert and S. Cohn-Vossen, Geometry and the Imagination, Chelsea, New York, 1952 (German original: Springer, Berlin, 1932).
4. N. H. Kuiper, Der Satz von Gauss-Bonnet und damit verwandte Probleme, Jber. Deutsch. Math.-Verein., 69 (1967) 77–88.
5. J. Milnor, Morse Theory, Princeton University Press, 1963.
6. G. Pólya, An elementary analogue of the Gauss-Bonnet theorem, this MONTHLY, 61 (1954) 601–603.

FUNDAMENTA MATHEMATICAE: AN EXAMINATION OF ITS FOUNDING AND SIGNIFICANCE

SISTER MARY GRACE KUZAWA, Holy Family College, Philadelphia

Fifty years ago in Poland, a small but determined group of mathematicians launched a daring project—a specialized mathematical journal, FUNDAMENTA MATHEMATICAE. Intuitively aware that the mathematical works of the late nineteenth, and the early years of the twentieth centuries, were but a forecast of tremendous and triumphant achievements to come, many scholars immediately recognized the value of the mathematical research incorporated in the pages of a journal.

One of the first scholars to grasp and to acknowledge the value of the journal issuing from Poland was the greatly respected editor of the AMERICAN MATHEMATICAL MONTHLY, Raymond Clare Archibald. As early as September 1921 he wrote: "Of the ten mathematical periodicals started since January 1919, none are of such notable importance for mathematical research as FUNDAMENTA MATHEMATICAE of which two volumes have been published: the first (224 pages) in 1920, the second (287 pages) in 1921, before May 1" [1].

Growing awareness of the Journal. Consequently, during the years from 1920 to 1939, the publication FUNDAMENTA MATHEMATICAE was widely read and generally acclaimed to be a major achievement by the mathematical experts. When at the end of 1935 the twenty-fifth volume of the journal came off the presses, both Polish and foreign scholars greeted the journal with considerable enthusiasm. The Polish newspaper, WIADOMOŚCI LITERACKIE (Literary News) hailed the appearance of this volume as a remarkable event, "heralding a holiday for Polish mathematics" [2]. THE BULLETIN OF THE AMERICAN MATHEMATICAL SOCIETY observed that "the twenty-fifth volume of FUNDAMENTA MATHEMATICAE represents a notable event in the mathematical life of the whole world" [3].

Throughout the ensuing years the AMERICAN MATHEMATICAL MONTHLY kept its readers informed about the research emanating from Poland either through comments or reviews of works published. It is fitting and proper, therefore, that today, fifty years and sixty-three volumes later, the MONTHLY presents for a more thoughtful consideration, a brief history of the journal. It is sincerely hoped that the following pages will reveal to the younger generation of mathematicians that FUNDAMENTA MATHEMATICAE was and is a unique periodical which attracted international recognition and cooperation, and that it had considerable influence on the development of the modern theory of functions and point sets.

Founder's concept of scope and content. In 1920, Zygmunt Janiszewski, Stefan Mazurkiewicz, and Wacław Sierpiński, professors of mathematics at the University of Warsaw, founded the publication. The idea of initiating a specialized periodical with articles written in French, English, German, and Italian and explicitly devoted to one or two branches of mathematics was due chiefly to the gifted mathematician, Zygmunt Janiszewski.

Nowhere is the originality and insight of Janiszewski more apparent than in the article of 1918, *On the Needs of Mathematics in Poland*, wherein this idea was first expressed. For with bold strokes of the pen, Janiszewski not only sketched a detailed plan for a school of mathematics but, what is more important, ensured its existence and success by proposing this daring and novel venture, a specialized publication. Deeply convinced of the journal's feasibility, he assured his countrymen in the article that a publication of this kind would be indispensable to mathematicians specializing in the branches covered by the journal and that not only would the work gain readers but also respected co-workers from abroad. Janiszewski underscored the fact that "the very existence and distribution of such a journal would assure the world of our cultural life" [4].

This account makes it obvious that Janiszewski envisioned a publication international in scope, one in which not only would articles be written in varied languages but to which foreign mathematicians would be encouraged to submit their original studies.

Appearance of first volume. The first volume of FUNDAMENTA MATHEMATICAE, however, deviated slightly from the proposed plan in that it carried articles written only by Polish mathematicians. This departure from the original objectives called attention to the fact that there was no lack of creative Polish mathematicians who were able not only to initiate such a specialized publication but also to ensure its continuance. Janiszewski planned this strategy. "My aim," he wrote, "is to introduce all the Polish mathematicians who specialize in the theory of sets in the very first volume of this publication" [5].

Janiszewski's untimely death on January 3, 1920, deprived him of enjoying the realization of the publication. But Stefan Mazurkiewicz and Wacław Sierpiński, who collaborated with Janiszewski in the project and who were imbued

with the same ideas and ideals as Janiszewski, completed, edited, and published the first volume of FUNDAMENTA MATHEMATICAE that year.

The first volume opened with a portrait frontispiece and a brief sketch of Janiszewski. Twenty-four articles dealing with the theory of sets and related questions by ten authors comprised the bulk of the publication. An added innovation expressly challenging the mathematicians, a set of ten original problems contributed by specialists in the field, concluded the 224-page journal.

Immediately upon publication, the journal succeeded in attracting the interest of mathematicians throughout the world. Some noted that "of the ten mathematical periodicals started since January, 1919, none are of such notable importance for mathematical research as FUNDAMENTA MATHEMATICAE" [6]. There were others, however, who greeted this venture with some scepticism; among them was a leading mathematician of the period, Henri Lebesgue. In a letter to Professor Sierpiński, Lebesgue praised the papers in the first volume but expressed his strong doubt as to the possibility of obtaining enough material to maintain such a highly specialized journal as a going concern [7].

Lebesgue and the Journal. Despite his doubts, Lebesgue nevertheless contributed his article *Sur les Correspondances entre les points de deux espaces* to the second volume of FUNDAMENTA MATHEMATICAE, published in April 1921. After the appearance of this volume, Lebesgue also wrote a twelve page article, *A Propos d'Une Nouvelle Revue Mathématique*: FUNDAMENTA MATHEMATICAE, for the *Bulletin des Sciences Mathématiques.*

Interesting to observe are some of the comments contained in Lebesgue's article. Here is how Lebesgue describes the emergence of the Polish School on the International scene:

> "In 1919, among the events which became known, is that there happens to exist in Warsaw a small group of mathematicians daring enough to venture to produce a novel mathematical periodical and sufficiently in love with their science to find time to work in a manner so successful that the two volumes of FUNDAMENTA now published are almost entirely filled with their own treatises" [8].

(Besides Lebesgue's article, the second volume carried an article by Hans Hahn of the University of Bonn, and one by Nicolas Lusin of the University of Moscow.)

Continuing the discussion of FUNDAMENTA MATHEMATICAE, Lebesgue quoted the editorial note, "that the journal is devoted to the theory of sets and the questions related to it; the immediate applications of the theory of sets, analysis situs, mathematical logic, the foundations of mathematics."

Journal's goal in concentration. For the sake of clarity and completeness of information it is necessary to digress here. Up to the year 1928, the editorial staff of FUNDAMENTA MATHEMATICAE consisted not only of Stefan Mazurkiewicz and Wacław Sierpiński but also of Jan Łukasiewicz, an internationally recognized authority on mathematical logic, and Stefan Leśniewski, who were to be responsible for the articles on mathematical logic and the foundations of mathematics. It was the inclusion of these last two mentioned fields into the

early plan of the publication that induced the founders to name the journal FUNDAMENTA MATHEMATICAE. Also, a particular manner of the presentation of materials in the journal was agreed upon: the first volume would be devoted exclusively to the theory of sets and the second would present articles treating mathematical logic and the foundations of mathematics. This alternating pattern was to be a characteristic feature of the FUNDAMENTA MATHEMATICAE. However, after much deliberation, the editors of the publication decided that there was yet an insufficient number of Polish mathematicians specializing in the second area. The original plan was abandoned. In the final layout, the theory of sets and questions relating to it—topology and the theory of the real functions—became the subjects of concentration for the journal.

More to the purpose of the study, however, is Lebesgue's shrewd observation that "the theory of sets has been placed outside of the field of mathematics long enough by the high priests of the theory of analytic functions" [9]. He further commented that "this ostracism is now abating due to the fact that the theory of sets which branched from the theory of analytic functions is proving itself useful to its older sister and is convincing people of good faith of its value and riches" [10].

From the preceding citations it can be deduced that in 1922, the date of Lebesgue's article, the theory of sets was still in its early stages. It is also clear that an early venture into the theory of sets by a publication such as FUNDAMENTA MATHEMATICAE would most likely have the opportunity to play some part in the development of the new mathematical discipline.

FUNDAMENTA MATHEMATICAE effect international reverberations. Perhaps it may still be too early to assess precisely the role played by the individual mathematicians in the enormous evolution occurring today in the mathematical sciences and being caused by the new methods associated with set theory and its related fields. But surely it is not too soon to assert that there is sufficient evidence that the important contributions to the above fields in the FUNDAMENTA MATHEMATICAE played an important role in the great upsurge of set theory and its related fields that began in the early twenties and is continuing today.

Some means of recognition of this fact is provided by the references to the FUNDAMENTA MATHEMATICAE found in many of the scholarly treatises on set theory and its allied fields. For example, in *Abstract Set Theory* (1953), Abraham Fraenkel, professor at the University of Jerusalem, listed in his bibliography over one hundred references from the FUNDAMENTA MATHEMATICAE. Professor Solomon Lefschetz of Princeton University referred to thirty-five articles appearing in various volumes of FUNDAMENTA MATHEMATICAE in his book *Topics in Topology*. Twenty-four references are alluded to in *Set Topology* (1960) by R. Vaidyanathaswamy, professor at the University of Madras, India, while the outstanding Dutch mathematician, Adriann C. Zaanen of the University of Amsterdam, includes eleven references from FUNDAMENTA MATHEMATICAE in his *Theory of Integration*.

Another detail of evidence of the role of FUNDAMENTA MATHEMATICAE in the progress of set theory and its related fields is the fact that not only did mathematicians of other countries draw upon the findings in the journal but they also contributed many original studies to the publication. Thus, the twenty-fifth volume of FUNDAMENTA MATHEMATICAE, which was twice as large as the other volumes (597 pages), had among its contributors: Alexandroff of Moscow, Borel of Paris, Zermelo of Freiburg, Hardy and Littlewood of Cambridge, Hausdorff of Bonn, Hurewicz of Amsterdam, Fraenkel of Jerusalem, Vietori of Innsbruck, Cech of Brno, Hille of Yale, von Neumann of Princeton, Moore of Austin, and others.

FUNDAMENTA MATHEMATICAE and Set Theory. More convincing and perhaps even greater evidence of the part played by FUNDAMENTA MATHEMATICAE in the upsurge of set theory and its related fields are the following facts.

Between the years 1920 to 1935, one hundred and seventy authors, two-thirds of whom were scholars of countries other than Poland, contributed 732 scholarly papers. Four years later, with the publication of the thirty-second volume (1939), the numbers increased to 972 works by two hundred and sixteen authors. This constituted an increase of 240 works and forty-six authors over the four-year period. Surely, this is an excellent indication of the growing interest in set theory and its related fields.

Once the interest of the creative mathematicians was aroused, from all parts of the world a seemingly endless, constant flow of excellent and scholarly articles on set theory and its related fields was directed to the offices of FUNDAMENTA MATHEMATICAE. As a result, during the summer of 1939 the thirty-third volume was under preparation. Much of the volume was already set up in type and the remaining articles were being processed. And then, on September 1, 1939, Germany invaded Poland.

War severs continuity, progress. World War II—history's most terrible conflict—began. The savage, six-year struggle was greatly detrimental to the mathematical progress and creativity in Poland.

The thirty-third volume of FUNDAMENTA MATHEMATICAE never came off the presses. The German director of the printery, Walter Hande de Dresde, had the plates destroyed and the galleys and remaining manuscripts burned. On September 1, 1942, fires which broke out all over Warsaw completely destroyed the offices of FUNDAMENTA MATHEMATICAE and the mathematical holdings at the University of Warsaw. Nothing remained of the archives and the first editions of FUNDAMENTA MATHEMATICAE. The methodical block-by-block destruction of the city of Warsaw from August to December of 1944 by the Brandkommando resulted in the loss not only of the private libraries of the Polish mathematicians living in Warsaw but also of their manuscripts containing the original results produced during the period 1939–1944.

Most damaging to the field of mathematics, however, were the vile and inhuman acts perpetrated by the Nazis. The systematic and brutal slaughter of the

inhabitants deprived Poland of some of her finest mathematicians. In 1940, Aleksander Rajchman perished in the concentration camp at Dachau; Stefan Kempisty died in a prison camp. Victims of a surprise dragnet, four outstanding mathematicians, Antoni Łomnicki, Stanisław Ruziewicz, Włodzimierz Stożek, and Julius P. Schauder were among the more than thirty Professors from the Lwów area shot by the ruthless Gestapo in July, 1941. November, 1942, marked the murder by the Nazi conquerors of Stanisław Saks, the well-known author of the mathematical classic, *Théorie de l'Intégrale.* The gas chamber at Treblinka (January 1943) claimed the life of Józef Zalcwasser [**11**].

Of the 100 prominent creative mathematicians, 62 were murdered or died from sickness, starvation, or disease in the enemy concentration camps [**12**]. There were also about 32 younger and promising mathematicians (who had not yet attained world renown) who perished during the war. Among the latter, Józef Marcinkiewicz, who died in a prison camp in 1940 at the age of thirty, was endowed with extraordinary mathematical talent. Examining some of the mathematical findings left by Marcinkiewicz, some mathematicians, among them Antoni Zygmund, claim that if Marcinkiewicz lived and continued to work he would surely have become one of the leading mathematicians of today [**13**].

Some mathematicians, fortunately, did survive the concentration camps. These relate that the study of mathematics continued in Poland in spite of hardships and unfavorable circumstances. One of those more active in teaching mathematics in the underground was Professor Kazimierz Zarankiewicz, who in 1944 was sentenced to hard labor in the heart of Germany. Prison walls could not and did not halt the imprisoned Professors from creating and exchanging ideas. Professor Stanisław Urbańczyk of the University of Cracow in his book, *Uniwersytet za Kolczastym Drutem* (*A University Behind Barbed Wires*), credits the survival of many of the intelligentsia to the intellectual life and cultural atmosphere which the professors had created for themselves. (The enemy was completely unaware of the lectures and discussions which kept up the morale of the war victims.) [**14**].

The coming of peace in 1945 found much of Poland reduced to rubble and ashes. The uncertainties of the postwar era and the fear of communism seemed to toll a death knell for the mathematical field in Poland. But, surprisingly, no sooner was peace declared, than FUNDAMENTA MATHEMATICAE appeared, containing articles just as creative and promising for further research as those found in the preceding volumes.

If credit is to be given the individuals who contributed most to this reactivation, the competence, courage, and devotion of the great Polish mathematicians, Wacław Sierpiński, Kazimierz Kuratowski, and Karol Borsuk must be recognized. As editors of the FUNDAMENTA MATHEMATICAE, they dedicated the thirty-third volume of FUNDAMENTA MATHEMATICAE to the memory of their colleagues, contributors to the journal, who perished during the war. On the same page of the journal the editorial staff announced the loss of one of their great mathe-

maticians, Stefan Banach, who died on August 31, 1945. Banach was greatly affected by the atrocities committed by the Nazis. When peace came, it found him a broken-hearted man for whom life no longer held a promise. Also of special note to the reader in the thirty-third volume are the 115 pages which contain work of the Polish mathematicians which evolved secretly during the war years.

The mathematical world did not expect FUNDAMENTA MATHEMATICAE in 1945. The sentiments of the mathematicians regarding the journal's continuance are best expressed by Professor Paul Aleksandrov of Russia. Upon receiving a copy of the publication, Aleksandrov wrote to the Polish mathematicians that he was greatly aggrieved by the unfortunate disasters that plagued the progress of mathematics in Poland. In fact, he wondered if the journal would survive the tragedy. Subsequently, he observed that this early and excellent publication of the periodical "delights me very much for it is a symbol of triumph of the eternal ideals of man's culture" [15]. Unquestionably the thirty-third volume proves a monument to the undaunted spirit that characterized and characterizes the Polish mathematicans.

Conclusion. Although the continuity of publication, the international scope of its author-contributors, and the quantity of articles representing a spectrum in mathematical research are most impressive, perhaps the greatest claim to its significance is the excellence of the scholarly works which gained worldwide recognition for FUNDAMENTA MATHEMATICAE. Some of the more outstanding treatises helped significantly in the development of the mathematical literature of the world. A few from among these are: Lebesgue's *Sur les correspondances entre les points de deux espaces* (Volume II), which laid the foundations for measure theory; Banach's *Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales* (Volume III), which initiated a new mathematical discipline, functional analysis: Urysohn's *Mémoire sur les multiplicités Cantoriennes* (Volumes VII and VIII), which presented all the fundamental results in measure theory; and Lusin's *Sur les ensembles analytiques* (Volume X), which embodied all the principal results of the theory of analytic functions [16].

Journal remained an unchanging clearinghouse. At this point it should also be noted that collaboration in mathematical research was always encouraged by the editors of FUNDAMENTA MATHEMATICAE. As a result, a great number of studies in the journal were and are the work of two or even three scholars, not necessarily from the same mathematical centers. For example, in the twenty-fifth volume, the article *Note on the differentiability of multiple integrals* was authored by B. Jessen of Copenhagen, J. Marcinkiewicz and A. Zygmund of Poland.

In concluding the discussion of FUNDAMENTA MATHEMATICAE, a final but important observation must be made. The excellent planning and the extraordinary cooperation of the members of the Polish School of Mathematicians produced a journal which in general retained its original format in the 63 volumes published.

Two changes that occurred were necessitated by growth. Commencing with 1929, studies on functional analysis appeared in the newly founded periodical, STUDIA MATHEMATICA, which became an outstanding publication in that field. The section *Problèmes* was assumed by the *Scottish Book*, a notebook totally devoted to problems. The book, placed in a Lwów café, was at the disposal of every mathematician who demanded it. Sharing Hilbert's view that the formulation of a problem is halfway to its solution, the Poles were not only fond of problems but were able to devise very challenging ones. In 1947, the Paris publication INTERMÉDIAIRE DES RECHERCHES MATHÉMATIQUES reprinted all the problems appearing in the thirty-two volumes published in FUNDAMENTA MATHEMATICAE. It added: "This publication, specializing in the theory of sets, has contributed very strongly to the progress of modern mathematics" [17].

In concluding, there is need to add that although so many great Polish mathematicians have perished during the war, mathematics is flourishing. At present in Poland there are 500 creative mathematicians who contribute on the average about 400 creative mathematical works per year to the nine mathematical journals [18]. There is no doubt that the Polish mathematicians contributed significantly and continue to contribute strongly to the progress of modern mathematics.

References

1. R. C. Archibald, Notes, this MONTHLY, 28 (1921) 317.
2. Edward Marczewski, Świeto Matematyki Polskiej, Wiadom. Literackie, 19 (1936) 6.
3. Jacob Tamarkin, Twenty-five volumes of Fundamenta Mathematicae, Bull. Amer. Math. Soc., 42 (1936) 300.
4. Zygmunt Janiszewski, O Potrzebach Matematyki w Polsce, Nauka Polska, 1 (1918) 20.
5. Kazimierz Kuratowski, Piećdziesiat Tomów Fundamenta Mathematicae, Wiadom. Mat., 7 (1963) 11.
6. Archibald, *op. cit.*, 317.
7. Kuratowski, *op. cit.*, 12.
8. Henri Lebesgue, A Propos d'une Nouvelle Revue Mathématique: Fundamenta Mathematicae, Bull. Sci. Math., 46 (1922) 35–48.
9. *Ibid.*, 36.
10. *Ibid.*
11. Kazimierz Kuratowski, La Rédaction, Fund. Math., 33 (1945) 9.
12. S. Hartman, Sprawozdanie Rozwoju Matematyki w Polsce, 1945–1963, Wiadom. Mat., Ser. 2, 8 (1965) 1.
13. Antoni Zygmund, Józef Marcinkiewicz, Wiadom. Mat., Ser. 2, 4 (1960) 11.
14. Stanisław Urbańczyk, Uniwersytet za Kolczastym Drutem, Kraków: Nakładem Ksiegarni Stefana Kamińskiego, (1946) 67.
15. Kuratowski, *op. cit.*, p. 17.
16. Kazimierz Kuratowski, Piećdziesiat Tomów Fundamenta Mathematicae, Wiadom. Mat., 7 (1963) 14–15.
17. Paul Belgodére (ed.), Intermédiaire des Recherches Mathématiques, 3 (1947) 3.
18. Hartman, *op. cit.*, p. 1.
19. Sister Mary Grace Kuzawa, Modern Mathematics: The Genesis of a School in Poland, College & University Press, New Haven, 1968, pp. 101–111.

THE JEEP ONCE MORE OR JEEPER BY THE DOZEN

DAVID GALE, University of California at Berkeley

1. Introduction. In 1947, N. J. Fine solved the by now famous problem of the jeep [1]. We recall that the problem concerns a jeep which is able to carry enough fuel to travel a distance d, but is required to cross a desert whose distance is greater than d (for example $2d$). It is to do this by carrying fuel from its home base and establishing fuel depots at various points along its route so that it can refuel as it moves further out. It is then required to cross the desert on the minimum possible amount of fuel.

Our purpose here is first to give a very short derivation of the solution of the jeep problem which makes use of a theorem (also famous) of Banach [2]. Second, we consider the situation in which it is required to send a jeep across the desert every day, say, for a week. Of course, having found the best procedure for a single jeep, one could simply repeat this seven times. We shall show, however, that there is a more economical procedure for the case of several jeeps, and in general that the more jeeps one sends across, the lower fuel consumption per jeep. This phenomenon is an example of what economists refer to as "increasing returns to scale," a subject of some economic interest. (It also accounts for the subtitle of this article, for which I apologize herewith.)

2. The problem. To formalize the problem, let us assume that the jeep starts from the origin and moves along the positive x-axis. We choose for the unit of fuel the maximum amount which the jeep can carry, and refer to this unit as a *load*. The unit of distance will then be chosen as the distance the jeep can travel on one load.

Figure 1 below gives a schematic representation of a typical jeep's journey.

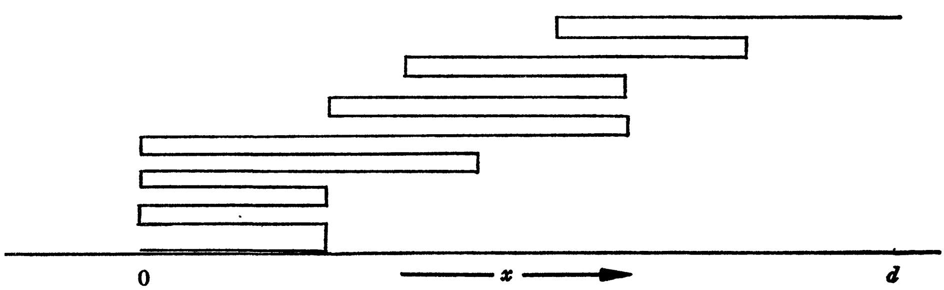


FIG. 1

The wiggly path represents the jeep's travel. Of course, actually the path lies entirely on the x-axis. It has been stretched vertically simply to make it visible. Note that because of our choice of units, the length of this path is precisely equal to the amount of fuel consumed. In the figure, the jeep reaches a point at dis-

David Gale is known for his research in Mathematical Economics, Game Theory, and the Geometry of Convex Sets. After taking his degree at Princeton under A. W. Tucker, he taught at Princeton, Brown, and (his present appointment) Berkeley. He spent a Fulbright year in Copenhagen, a Guggenheim year in Osaka, was a Berkeley Visiting Miller Professor another year, and currently is back in Copenhagen as an NSF Fellow. He is a Fellow of the Econometric Society. *Editor.*

tance d from the origin. The original jeep problem asks for the minimum fuel consumption (hence path length), which will allow the jeep to reach this point.

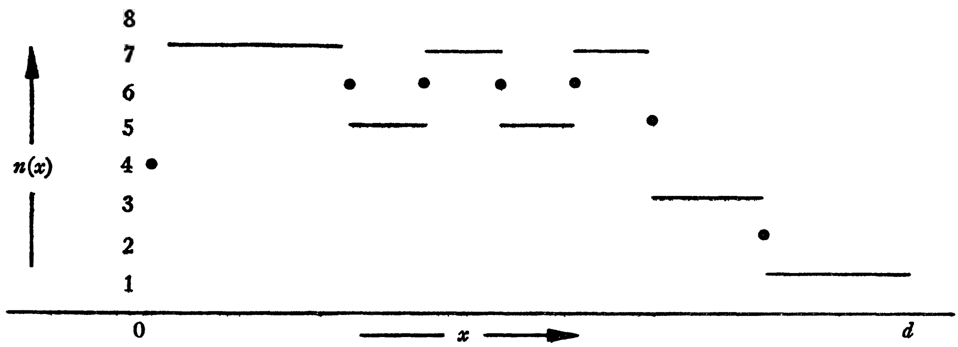


FIG. 2

It turns out to be somewhat more convenient to turn the problem around (see e.g. [3]) and obtain a formula for the function $d(f)$ giving the farthest point which the jeep can reach on f loads of fuel. Our first aim is to prove the formula

$$d(f) = 1 + \frac{1}{3} + \frac{1}{5} + \cdots + \frac{1}{2f-1}\,. \tag{1}$$

The key idea in our solution is to make use of a formula of Banach for the path length of a curve in one dimensional space. (Path length in one-space is usually referred to as *total variation*; we prefer the geometric terminology as being more suggestive in the present context.) To utilize Banach's Formula we define for each point x on the interval $[0, d]$, the *valance* $n(x)$ as the number of times during its journey that the jeep is at the point x. Figure 2 above gives the graph of the valance $n(x)$ corresponding to the jeep's journey plotted in Figure 1. Strictly speaking, if one allowed very general paths, $n(x)$ might be infinite for some points. This would not affect Banach's Formula, which states

$$\text{total path length} = \int_0^d n(x)dx.$$

Of course, Banach was not concerned with jeeps. He considered a continuous real valued mapping $x=\phi(t)$, and $n(x)$ was simply the number of points which mapped onto the point x, or the cardinality of $\phi^{-1}(x)$. We can also think of the problem this way if we take the wiggly line of Figure 1 to be the graph of the jeep's position plotted, say, as a function of the time.

For "reasonable" paths, Banach's Formula is obvious. A reasonable path for a jeep is one with a finite number of points at which it reverses direction (it would be a remarkable jeep indeed that could execute an unreasonable path). To prove the formula, partition $[0, d]$ into sets $X_1, X_2, \cdots$, where

$$X_k = \{x \mid n(x) = k\}.$$

Because of reasonableness, there are only finitely many nonempty X_k, and each of them is a union of disjoint intervals. (Some of them may consist of single points. In the case of the jeep's tour, this will be the case whenever k is even. Do you see why?) Over each interval of X_k lie exactly k intervals of the jeep's path. Therefore,

$$\text{total path length} = \sum_{k=1}^{\infty} k \text{ (length of } X_k).$$

But the term on the right is precisely the definition of $\int_0^a n(x)dx$. (Notice that this is the definition of the integral in the sense of Lebesgue rather than Riemann! Of course for continuous functions, and in particular for reasonable functions, the two concepts are equivalent.) We remark that Banach's formula holds for unreasonable as well as reasonable paths (though the general proof is fairly involved), and hence our solution of the jeep problem will hold for unreasonable as well as reasonable jeeps.

3. The solution for one jeep. We return now to the problem of computing $d(f)$, and assume for the moment that f is an integer. We wish to determine how far the jeep can get on f loads of fuel. For any jeep's tour, we define the sequence of points $x_0, x_1, \cdots, x_f$ on the interval $[0, d]$ where $x_f = 0$, $x_0 = d$, and in general x_k is the point such that the total path length (hence fuel consumption) to the right of x_k is exactly k units. Clearly the points x_k form a strictly decreasing sequence and there will be exactly one unit of path length between x_{k+1} and x_k. The basic observation we need is the following:

LEMMA 1. *If $x < x_k$, then*

$$n(x) \geqq 2k + 1. \tag{2}$$

Proof. Since x is to the left of x_k, the jeep must consume more than k loads of fuel to the right of x. Since the jeep can only carry one load at a time, it must therefore cross the point x at least $k+1$ times from the left. But between any two crossings from the left there must be a crossing from the right, so there must be at least k crossings from the right. Then the jeep must arrive at the point x at least $2k+1$ times, which is what the lemma asserts.

We now combine this result with Banach's Formula and get

$$\begin{aligned} 1 &= (\text{path length between } x_{k+1} \text{ and } x_k) \\ &= \int_{x_{k+1}}^{x_k} n(x)dx \geqq (2k+1)(x_k - x_{k+1}) \end{aligned}$$

so $x_k - x_{k+1} \leqq 1/(2k+1)$. Summing from 0 to $f-1$ implies

$$\sum_{k=0}^{f-1} (x_k - x_{k+1}) = x_0 - x_f = d \leqq 1 + \frac{1}{3} + \frac{1}{5} + \cdots + \frac{1}{2f-1}, \tag{3}$$

which gives an upper bound for $d(f)$. It remains only to show that this bound can be achieved, and this is easily done by induction. The formula is clearly correct when $f=1$. Suppose now we are allowed $f+1$ loads. Then let the jeep take $f+1$ loads to the point $1/(2f+1)$. This will involve $f+1$ outward trips and f return trips, hence $2f+1$ trips of length $1/(2f+1)$; therefore a total of one load will be consumed leaving f loads deposited at the point $1/(2f+1)$. By the induction hypothesis, the formula of (1) holds from this point on, completing the proof.

For the case where f is not an integer, the same type of argument shows that

$$d(f) = 1 + \frac{1}{3} + \cdots + \frac{1}{2[f]-1} + \frac{\{f\}}{2[f]+1}, \tag{4}$$

where $[f]$ and $\{f\}$ are the integral and fractional part of f respectively. In other words, one simply interpolates linearly between integral values of f. The graph of $d(f)$ is plotted in Figure 3.

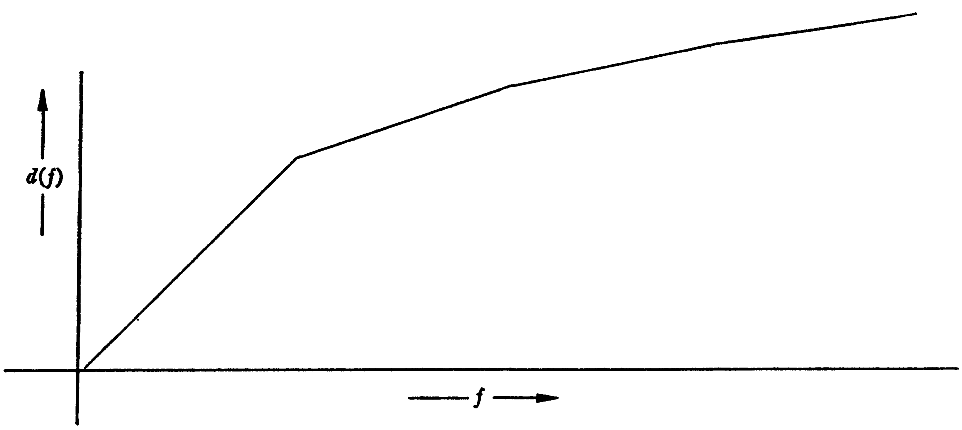


FIG. 3

Since the odd harmonic series diverges, it follows that a desert of any size can be crossed.

I should remark that after finding the solution just given, I became aware of the rather similar one given in [4]. The use here of the Banach Formula seems to tighten the argument somewhat and make the reasoning more transparent. Also we do not need to assume *a priori* that there will be only a finite number of depots. It is at least conceivable that the optimal solution would involve infinitely many deposits, or even a sort of continuous smear of fuel spread out along the route. One could no doubt formulate a very general problem in terms of measure theory. The argument above shows, however—thanks again to Banach's Formula—that this more general behavior could not give any improvement in fuel consumption.

Before leaving the single jeep, we consider the case in which the jeep is required to cross the desert and then return. For this case the arguments are the

same as before except that (2) becomes

$$n(x) \geqq 2k + 2. \tag{5}$$

Letting $\tilde{d}(f)$ be the longest possible round trip (e.g., twice the distance to the farthest point), we get the even simpler formula

$$\tilde{d}(f) = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{f}, \tag{6}$$

where as before equality can be achieved. We note the familiar fact that a round trip can be substantially cheaper than two one way trips. In fact comparing (6) with (1), we have

$$d(f) - \frac{1}{2}\tilde{d}(f) = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots - \frac{1}{2f};$$

but this is bounded by $\sum_{n=1}^{\infty} (-1)^{n+1}(1/n) = \log 2$. Thus for long distance, the increase in fuel cost for a round trip as against a one way trip becomes negligible.

4. Several jeeps. We turn now to the problem of m jeeps, and we shall compute the function $d_m(f)$ giving the most distant point which all jeeps can reach if they have f loads of fuel to share between them. In other words, the sum of the path lengths of all of the jeeps must not exceed f.

We proceed by defining the points x_k exactly as in the one jeep case, as the point to the right of which the total combined path length of all jeeps is k. The analogue of Lemma 1, however, is slightly more complicated.

LEMMA 2. *For any x in $[0, d]$, $n(x) \geqq m$. If $x < x_{m+r}$ $(r \geqq 0)$ then*

$$n(x) \geqq m + 2r + 2. \tag{7}$$

Proof. The first assertion simply corresponds to the fact that all m jeeps are required to reach the point d. Concerning inequality (7), we see that more than $m+r$ loads must be transported beyond the point x, so this point must be crossed at least $m+r+1$ times from the left. But since there are only m jeeps, in order to achieve $m+r+1$ crossings from the left there must be at least $r+1$ crossings from the right, giving $m+2r+2$ crossings in all.

We proceed to calculate $d_m(f)$. Notice that for $f \leqq m$, the problem is trivial and $d_m(f) = m/f$ (why?) so we assume that $f = m+s$, where $s > 0$. Again restricting ourselves to integral values of s, we claim

$$\begin{aligned} d_m(f) &= 1 + \frac{1}{m+2} + \frac{1}{m+4} + \cdots + \frac{1}{m+2s} \\ &= 1 + \frac{1}{m+2} + \frac{1}{m+4} + \cdots + \frac{1}{2f-m}, \end{aligned} \tag{8}$$

for the path length from x_{k+1} to x_k is one, so

$$1 = \int_{x_{k+1}}^{x_k} n(x)dx \geqq m(x_k - x_{k+1}), \qquad \text{for } k < m,$$

and

$$1 = \int_{x_{m+r+1}}^{x_{m+r}} n(x)dx \geqq (m + 2r + 2)(x_{m+r} - x_{m+r+1}).$$

Therefore $x_k - x_{k+1} \leqq 1/m$ for $k < m$, and

$$x_{m+r} - x_{m+r+1} \leqq \frac{1}{m + 2r + 2} \qquad \text{for } r \geqq 0.$$

Summing for $0 \leqq k \leqq m-1$ and $0 \leqq r \leqq s-1$ gives (8) as an inequality. Again an inductive proof shows that equality can be achieved. Assuming the formula correct for $m+s$ loads of fuel, we see that $m+s+1$ loads can be moved to the point $1/(m+2s+2)$ by having one jeep make $s+1$ round trips and then having all m of them make the one way trip to this point. This will use up one load, so that $m+s$ loads are deposited, and the induction can be continued. Note that there are various ways in which this optimal journey can be performed. One way is to have one of the jeeps do all the work of setting up the various depots, while the others simply move outward without turning around, refueling as they go. It is, however, not possible to prescribe the same path for all jeeps, for if all jeeps followed the same route, then the function $n(x)$ would have to be divisible by m at all points, which it clearly is not in an optimal trip. We shall return to this point later.

To prove the result about increasing returns, we wish to compare the distance which one jeep goes on f loads with that which m jeeps go on mf loads. For this purpose we rewrite (8) for the case $s = m(f-1)$ as follows:

$$\begin{aligned} d_m(mf) = 1 &+ \left(\frac{1}{m+2} + \cdots + \frac{1}{3m}\right) + \left(\frac{1}{3m+2} + \cdots + \frac{1}{5m}\right) + \cdots \\ &+ \left(\frac{1}{2(f-3)m+2} + \cdots + \frac{1}{(2f-1)m}\right), \end{aligned} \tag{9}$$

where each term in parentheses contains m summands and there are f terms in all. We may also rewrite (3) as

$$\begin{aligned} d(f) = 1 &+ \left(\frac{1}{3m} + \frac{1}{3m} + \cdots + \frac{1}{3m}\right) + \left(\frac{1}{5m} + \cdots + \frac{1}{5m}\right) + \cdots \\ &+ \left(\frac{1}{(2f-1)m} + \cdots + \frac{1}{(2f-1)m}\right), \end{aligned} \tag{10}$$

where there are also m summands in each term in parentheses. A term by term comparison between (9) and (10) shows the advantage of using many jeeps.

For the special case $m=2$, we get

$$d_2(2f) - d(f) = \left(\frac{1}{4} - \frac{1}{6}\right) + \left(\frac{1}{8} - \frac{1}{10}\right) + \cdots + \left(\frac{1}{4f-4} - \frac{1}{4f-2}\right)$$
$$= \frac{1}{2}\left(\frac{1}{2} - \frac{1}{3} + \frac{1}{4} - \frac{1}{5} + \cdots + \frac{1}{2f-2} - \frac{1}{2f-1}\right).$$

Thus the longer the trip the greater the saving. On the other hand since the terms in parentheses above are part of a convergent series, it follows that the total amount which is saved remains bounded as the trip gets longer.

For the round trip problem, the formula is again simpler. Let $\tilde{d}_m(f)$ be the round trip distance which m jeeps can achieve on f loads. Then

$$\tilde{d}_m(mf) = 1 + \frac{1}{m+1} + \frac{1}{m+2} + \cdots + \frac{1}{mf} \cdot \tag{11}$$

So again, the more jeeps there are, the further out they can get on the same amount of fuel; or returning to the original problem, the less fuel per jeep is needed to reach a preassigned point. But there is a limit to the advantage one obtains in increasing the number of jeeps; for as m approaches infinity, a standard calculation (comparison with $\int dx/x$) shows

$$\lim_{m\to\infty} \tilde{d}_m(mf) = 1 + \log f, \tag{12}$$

so no matter how many jeeps there are, one cannot go further than $1+\log f$ on f loads of fuel per jeep. In terms of the original problem, we get in the limit

$$f = e^{d-1} \tag{13}$$

so the amount of fuel needed increases exponentially with the length of the desert, as one might have guessed.

The multi-jeep round trip problem can be interpreted in another way. Instead of thinking of m jeeps, each making a round trip, one may consider a single jeep which must make round trips, say on m successive days. The solution here is the same as for the m jeep problem with the consequent saving of fuel. In fact, on the first day the jeep can set up all the depots it will need for the following days, as one easily sees.

5. Some Final Remarks and Questions. The last paragraph above raises an interesting question. Suppose instead of being concerned about m round trips across the desert, one has decided to go into the desert crossing business and plans to make a round trip daily into the indefinite future. What sort of routine should one then use? As we have just seen, it would be uneconomical to use the single round trip routine each day. Similarly, repeated use of the m-day routine would be inferior to using the $(m+r)$-day routine, so that no periodic program of this sort could be optimal. On the other hand, if one is to consider programs

which are not periodic, it is no longer clear what one should mean by an optimal program. In any case, it appears that the problem of finding the best "steady state" routine has no exact solution at all, so that in practice one would have to settle for a routine that was "almost optimal."

There are many other jeep problems one can think of. Helmer [5] has considered some fairly complicated situations in which the number of depots one is allowed to establish is limited. To get a feeling for this sort of problem, the reader might look at the problem of crossing a desert of length 2 when only 3 intermediate depots are permitted.

An apparently simple question is the round trip problem in which fuel is available at both ends of the desert, but I must confess with embarrassment that I have not been able to find the solution. It is not hard to see that one can do at least as well in this case as in the case of two jeeps making one-way trips, but it may be possible to do better. The difficulty here as with many optimization problems is that there does not appear to be any simple way to determine whether or not a given solution is optimal. The upper bound given by Banach's Formula does not seem to be available for this case. I put this problem forward as a challenge to jeepologists.

I conclude with some historical remarks. Shortly after the publication of Fine's solution, Phipps [3] derived the same result by arguing that the single jeep problem is equivalent to a problem involving a convoy of jeeps which travel together, some being used to refuel others. The solution to the convoy problem is very simple, but the argument that this problem is equivalent to the original problem does not seem to be quite complete. Using this equivalence, however, one can also easily derive the result given here on increasing returns. Finally, there seems to be a feeling among many people that the jeep problem can be solved by the functional equation method of dynamic programming. In fact the problem occurs as an exercise in the book of Bellman [6], but the solution is not given there and I know of no way of solving the problem by this method.

6. Addendum. It has been pointed out to me that if one accepts the convoy equivalence of Phipps, then dynamic programming can be used (see e.g. [7]). However, for the convoy problem, the solution is almost obvious anyway. Imagine that f jeeps set out. Since all but the last must return home, we suppose that $f-1$ of them consume fuel at the rate 2 and the last one, the "red, white, and blue" jeep that will make the final crossing, consumes at the rate 1. Thus, initially, the convoy is consuming at the rate $2f-1$, which it does until one load has been consumed, after which the first jeep can be abandoned. The remainder then go on consuming at the rate $2f-3$, etc.

I should also remark that for Helmer's variation in which the number of depots is specified, dynamic programming does seem to be the appropriate tool to use.

References

1. N. J. Fine, The jeep problem, this MONTHLY, 54 (1947) 24–31.

2. S. Banach, Sur les lignes rectifiables et les surfaces dont l'aire est finie, Fundamenta Mathematicae, VII (1925) 225–236.

3. C. G. Phipps, The jeep problem, A more general solution, this MONTHLY, 54 (1947) 458–462.

4. G. C. Alway, Crossing the desert, Math. Gazette, 41 (1957) 209.

5. O. Helmer, A problem in logistics. The jeep problem, Project Rand Report No. Ra 15015, Dec. 1947.

6. R. Bellman, Dynamic Programming, Princeton University Press, 1955, p. 103 ex. 54–55.

7. J. N. Franklin, The range of a fleet of aircraft, J. Soc. Indust. Appl. Math., 8 (1960) 541–548.

MATHEMATICAL NOTES

EDITED BY DAVID DRASIN

Manuscripts for this Department should be sent to David Drasin, Division of Mathematical Sciences, Purdue University, Lafayette, IN 47907.

A NOTE ON SECOND CATEGORY TOPOLOGICAL GROUPS

E. J. HOWARD, San Diego State College

In this note we prove that if G is a topological group and H is the closure of the identity, then G is of second category if and only if G/H is of second category. Consequently since G/H is a Hausdorff group the study of second category topological groups can in most cases be restricted to Hausdorff topological groups.

Let X be a topological space. A subset A of X is *nowhere dense* in X if the interior of $\overline{A}$ (the closure of A) is empty. A subset A of X is said to be of *first category* in X if it is the union of a countable family of nowhere dense sets. If a subset A of X is not of first category, then it is said to be of *second category* in X. Throughout this paper the topology of G/H will be the quotient topology.

We begin by giving some well-known facts concerning topological groups.

LEMMA 1. *Let G be a topological group and H a normal subgroup of G and $\phi: G\to G/H$ the natural homomorphism, then:*

a. *The mapping $\phi: G\to G/H$ is a continuous and open mapping.*

b. *If H is compact, then $\phi: G\to G/H$ is a closed mapping.*

c. *If e is the identity of G and $H=\overline{\{e\}}$, then H is a compact closed normal subgroup of G and G/H is Hausdorff.*

Proof. The following page references are all in [1]. For proofs of (a) and (b) see pages 36 and 37. For (c), that H is a closed normal subgroup is on page 33 and that G/H is Hausdorff is on page 45. By Proposition 3, Section 20 of [3]

$$H = \overline{\{e\}} = \cap U_\alpha, \quad \text{where} \quad \{U_\alpha\}$$

is the collection of all neighborhoods of e, thus for any open cover of H there is an element U in the cover containing e, hence $\bigcap U_\alpha \subseteq U$. Consequently H is compact.

LEMMA 2. *Let G be a topological group and $H=\overline{\{e\}}$ and $\phi: G\to G/H$, the natural homomorphism. If C is a closed subset in G and $C^0=\emptyset$ (C^0 denotes the interior of C), then $[\phi(C)]^0=\emptyset$.*

Proof. Suppose C is closed in G, $C^0=\emptyset$ and $[\phi(C)]^0\neq\emptyset$, then there exists a nonempty open set V in G/H such that $V\subseteq\phi(C)$ and an open neighborhood U of the identity in G such that for some x in G, $\phi(xU)\subseteq V$. Hence $xUH\subseteq CH$. The set $xUH=W$ is an open subset of G. We shall show that $CH\subseteq C$. Let $y\in CH$, then $y=ch$, where $c\in C$ and $h\in H$. Now by (5.4) of [1] and the fact that H is a subgroup of G we have $\overline{\{ch\}}=(ch)H=cH=\overline{\{c\}}$. Since $c\in C$ and C is closed, $\{ch\}\subseteq C$ and hence $ch\in C$. Thus the nonempty open set W is contained in C contrary to the fact that $C^0=\emptyset$.

Our main theorem is a generalization of a theorem in [2].

THEOREM. *Let G be a topological group and $H=\overline{\{e\}}$; then G is of second category if and only if G/H is of second category.*

Proof. We first show that if G is of second category and $G/H=\bigcup E_n$, then for some n, $\overline{E}_n^0\neq\emptyset$. Consequently, G/H is also of second category. To see this we observe that $G=\bigcup\phi^{-1}(E_n)=\bigcup\phi^{-1}(\overline{E}_n)$, and because ϕ is continuous each $\phi^{-1}(\overline{E}_n)$ is closed. Since G is of second category, for some n, $\phi^{-1}(\overline{E}_n)$ contains a nonempty open set, and because ϕ is open, $\overline{E}_n$ also contains a nonempty open set, i.e., $\overline{E}_n^0\neq\emptyset$. (Note: this proof holds for any normal subgroup of G.)

Now for the converse assume G is of first category; then $G=\bigcup F_n$, where $\{F_n\}$ is a countable collection of nowhere dense sets. By Lemma 1, for all n, $\phi(F_n)$ is closed and by Lemma 2 has empty interior. Thus G/H is a countable union of nowhere dense sets. Hence G of first category implies G/H is of first category. Consequently G/H of second category implies G is of second category.

References

1. E. Hewitt and K. Ross, Abstract Harmonic Analysis, Academic Press, New York, 1963.
2. E. Howard, First and second category Abelian groups with the n-adic topology, Pacific J. Math., 16 (1966) 323–329.
3. T. Husain, Introduction to Topological Groups, Saunders, Philadelphia, 1966.

ON THE SETS OF DIRECTIONS DETERMINED BY n POINTS

P. R. SCOTT, University of Adelaide

1. Introduction. Some years ago P. Erdös [1] raised the problem of finding the least number of different distances determined by n points in the plane. We consider here the somewhat analogous problem of finding the least number of different directions (gradients) assumed by lines joining n noncollinear points in the plane. Some generalizations of the problem are also given.

2. The function k_n. Let P_n be a set of n noncollinear points in the plane together with the set of all straight *connecting* lines which join these points. Let $k(P_n)$ be the number of different directions assumed by the connecting lines of P_n. If now $\mathcal{P}$ denotes the family of all such planar sets P_n, let

$$k_n = \min_{P_n \in \mathcal{P}} k(P_n).$$

It is clear that $k_3 = 3$ (for any triangle), $k_4 = 4$, $k_5 = 5$. In the following theorem we give bounds for k_n.

THEOREM 1. *The minimum number k_n of directions determined by a set of $\mathcal{P}$ satisfies the inequalities*

$$\tfrac{1}{2}\{1 + \sqrt{8n - 7}\} \leqq k_n \leqq 2[n/2].$$

Proof. The upper bound is easily obtained. If n is even, we take the n points as vertices of a regular n-gon. In this case, the sides and diagonals of the regular n-gon lie along sets of parallel lines. There are

(a) $n/2$ distinct directions determined by sets of parallels containing pairs of opposite sides of the n-gon, and

(b) $n/2$ distinct directions, each determined by a set of parallels perpendicular to a diagonal joining an opposite pair of vertices of the n-gon.

Further, if for each set of parallels in (b) we draw representative parallel lines through the given opposite vertices, it is readily seen that no line of (b) is parallel to a line of (a). Hence n directions are determined. For odd n, we simply add the center point; this addition leaves the number of determined directions unchanged.

We now obtain the lower bound. Suppose that P_n^* is a set of $\mathcal{P}$ for which the value k_n is attained. Let $Q \in P_n^*$ be an extreme (corner) point of the convex hull of P_n^*, and let $l_1, \cdots, l_r$ be the connecting lines joining Q to the remaining points of P_n^*. If l_i contains u_i points of the set $P_n^* (1 \leqq i \leqq r)$, we can easily show that the points on each pair of lines l_j, l_k determine at least $u_j + u_k - 1$ directions. For let R_j, R_k be the points closest to Q on l_j, l_k respectively. Joining R_j to all points on l_k, and R_k to all points on l_j, gives a set of $u_j + u_k - 1$ pairwise intersecting lines (the line R_jR_k occurring twice). Hence at least $u_j + u_k - 1$ directions are determined, and so

$$u_j + u_k \leqq k_n + 1 \qquad (1 \leqq j < k \leqq r).$$

Summing over all values of j, k we obtain

$$(r - 1)\sum_{i=1}^{r} u_i \leqq \frac{r(r-1)}{2}(k_n + 1); \text{ i.e., } \sum_{i=1}^{r} u_i \leqq \frac{r}{2}(k_n + 1).$$

If Σu_i is replaced by $n + r - 1$ (since Q has been counted r times) we then get

$$k_n \geqq \frac{2(n + r - 1)}{r} - 1.$$

The expression on the right takes its minimal value when r is as large as possible. Clearly for P_n^*, $r \leqq k_n$; hence

$$k_n \geqq \tfrac{1}{2}\{1 + \sqrt{8n - 7}\},$$

and the proof is complete.

Experiment would seem to indicate that $k_n = 2[n/2]$, but I have been unable to prove this.

3. k_n for convex sets. We recall that a set of n points is said to be *convex* if none of the given points lies in the interior of the convex hull of the points.

A complete result for convex sets is given by

THEOREM 2. *The minimum number of directions determined by a convex set of $\mathcal{P}$ is $k_n = n$.*

Proof. As before, let P_n^* be a (convex) set of $\mathcal{P}$ for which the value k_n is attained, and let Q be an extreme point of P_n^*. Using our previous notation, if l_1 and l_r are sides of the convex hull meeting at Q, then there are $n - u_1 - u_r + 3$ connecting lines through Q. Also, there are at least $u_1 + u_r - 3$ further directions determined by the remaining points on l_1 and l_r. Hence $k_n \geqq n$.

For even n, this result coupled with the upper bound of Theorem 1 shows that $k_n = n$. For odd values of n, the same conclusion follows by considering the regular n-gon, a figure which determines exactly n directions.

4. Extensions and generalizations. The problem has an obvious projective interpretation. Let S_n be a set of n noncollinear points in the projective plane together with all lines joining these points. We ask 'What is the smallest number, $k(S_n)$, of intercepts of the lines of S_n by a general line in the plane?' It is clear that $k(S_n)$ is not less than the k_n of Theorem 1.

We may also pose the original problem for nonplanar sets P_n. We then obtain

THEOREM 1′. *The minimum number k_n of directions determined by a nonplanar set of $\mathcal{P}$ in 3-space satisfies the inequalities*

$$\sqrt{6n - 6} \leqq k_n \leqq 2n - 2.$$

THEOREM 2′. *The minimum number k_n of directions determined by a strictly convex nonplanar set of $\mathcal{P}$ in 3-space satisfies the inequalities*

$$n + 2 \leqq k_n \leqq 2n - 2.$$

Proof. The lower bounds are obtained using methods exactly similar to those used in the planar case. The lower bound in Theorem 2′ is best possible for small n (for example the vertices of a tetrahedron), but is probably a poor bound for large n.

A figure for which the upper bound is attained is the pyramid having a regular $(n-1)$-gon as its base. The bound can be slightly improved to $2n-3$ for even n, by placing 2 suitable pyramids on opposite sides of a common regular $(n-2)$-gon base.

Reference

1. P. Erdös, On sets of distances of n points, this MONTHLY, 53 (1946) 248–250.

RESEARCH PROBLEMS

EDITED BY VICTOR KLEE

In this Department the Monthly presents easily stated research problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Material should be sent to Victor Klee, Department of Mathematics, University of Washington, Seattle, WA 98105.

A CONJECTURED REPRESENTATION OF GENOCCHI NUMBERS

J. M. GANDHI, Western Illinois University

Genocchi numbers are defined by the formula

$$(G+1)^N + G_N = 1, \qquad N > 1, \quad \text{with} \quad G_1 = 1, \tag{1}$$

where after expansion G^i is to be replaced by G_i for each $i \leqq N$. We remark that the Genocchi numbers can also be generated by the formula [1, pp. 250–263]:

$$\frac{2t}{e^t+1} = \sum_{N=0}^{\infty} \frac{G_N}{N!} t^N. \tag{2}$$

The first few Genocchi numbers are $G_1 = +1$, $G_2 = -1$, $G_4 = +1$, $G_6 = -3$, $G_8 = +17$, $G_{10} = -155$, $G_{12} = +2073$, $G_{14} = -38227$, $G_{16} = +929569$, etc. with $G_{2N+1} = 0$. We conjecture that

$$G_{2N} = (-1)^N \sum 1^2 \sum 2^2 \sum 3^2 \cdots \sum (N-1)^2, \tag{3}$$

where the $\sum$ notation used in (3) has the following meaning:

$$\begin{aligned} \sum k^2 &= k^2 - (k-1)^2 \\ \sum k^2 \sum (k+1)^2 &= k^2 \sum (k+1)^2 - (k-1)^2 \sum k^2 \\ &= k^2\{(k+1)^2 - k^2\} - (k-1)^2\{k^2 - (k-1)^2\} \end{aligned}$$

and in general we have the recurrence

$$\begin{aligned} &\sum k^2 \sum (k+1)^2 \sum (k+2)^2 \cdots \sum (k+N)^2 \\ &\qquad = k^2 \sum (k+1)^2 \sum (k+2)^2 \cdots \sum (k+N)^2 \\ &\qquad\quad - (k-1)^2 \sum k^2 \sum (k+1)^2 \cdots \sum (k+N-1)^2. \end{aligned} \tag{4}$$

We note that $(N+1)$ $\sum$'s on the left hand side of (4) are reduced to N $\sum$'s and the process can be continued till there are no $\sum$'s left. The $\sum$ notation used

can be easily understood by actually calculating the first few Genocchi numbers:

$$G_4 = \sum 1^2 = 1.$$
$$G_6 = -\sum 1^2 \sum 2^2 = -1^2 \sum 2^2 + 0^2 \sum 1^2 = -1^2(2^2 - 1^2) = -3.$$
$$G_8 = \sum 1^2 \sum 2^2 \sum 3^2 = 1^2 \sum 2^2 \sum 3^2 = 1^2(2^2 \sum 3^2 - 1^2 \sum 2^2)$$
$$= 1^2[2^2(3^2 - 2^2) - 1^2(2^2 - 1^2)] = +17.$$
$$G_{10} = -\sum 1^2 \sum 2^2 \sum 3^2 \sum 4^2$$
$$= -1^2[2^2(3^2 \sum 4^2 - 2^2 \sum 3^2) - 1^2(2^2 \sum 3^2 - 1^2 \sum 2^2)];$$

using (5) and simplifying we get $G_{10} = -155$.

The formula (3) has been verified to be true for all values of G's up to G_{14}.

For similar $\sum$ notation as used in this paper, though slightly different, reference may be made to [2].

References

1. Anatole Lucas, Théorie des Nombres, Vol. 1., Gauthier Villars, Paris, 1891.

2. C. Krishanamachary and R. M. Bhimsenrao, On a table for calculating Eulerian numbers based on a new method, Proc. London Math. Soc., Ser. 2, Vol. 22. (1921) 73–80.

CLASSROOM NOTES

EDITED BY DAVID DRASIN

Manuscripts for this Department should be sent to David Drasin, Division of Mathematical Sciences, Purdue University, Lafayette, IN 47907.

INTEGRATION OF TOTAL DIFFERENTIAL EQUATIONS

B. FISHER, The University, Leicester, England

Consider the total differential equation

$$P(x, y, z)dx + Q(x, y, z)dy + R(x, y, z)dz = 0, \tag{1}$$

where we shall suppose first of all that

$$Q(x, y, z) = P(y, z, x) \quad \text{and} \quad R(x, y, z) = P(z, x, y).$$

By treating x in (1) as a constant and integrating, we get

$$U(x, y, z) = \text{const.} = f(x). \tag{2}$$

Assuming (1) is integrable, there exists an f such that (2) is a solution of (1). The problem is to find f.

To do this, we have, by the symmetry of (1),

$$U(y, z, x) = f(y) \tag{3}$$

also is a solution of (1). Solving (3) for z, we have $z=V(x,y,f(y))$, and substituting this value of z in (2), we get $U(x,y,V)=f(x)$. By giving y some suitable fixed value α, then $f(\alpha)$ becomes some arbitrary constant c, and we get

$$f(x) = U(x, \alpha, V(x, \alpha, c)),$$

giving us the function f.

One must of course check that (1) is integrable before using this method, since the method of obtaining an f works even if (1) is not integrable.

Example: $yz(y+z)dx+zx(z+x)dy+xy(x+y)dz=0$. We have

$$\begin{vmatrix} yz(y+z) & zx(z+x) & xy(x+y) \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ yz(y+z) & zx(z+x) & xy(x+y) \end{vmatrix} = 0,$$

so the equation is integrable.

Treating x as a constant, we have $z(z+x)dy+y(x+y)dz=0$, and integration gives

$$\frac{(x+y)(x+z)}{yz} = f(x). \tag{4}$$

By symmetry, we have $[(y+z)(y+x)]/zx=f(y)$. Putting $y=1$ and solving for z, we get

$$z = \frac{x+1}{(c_1-1)x-1} = \frac{x+1}{cx-1},$$

where $f(1)=c_1$ is an arbitrary constant and $c_1-1=c$.

Substituting in (4) with $y=1$, we get

$$f(x) = \left\{x + \frac{x+1}{cx-1}\right\}(cx-1) = cx^2+1.$$

Thus the solution is $[(x+y)(x+z)]/yz=cx^2+1$, which can be rewritten to give the symmetric form $x+y+z=cxyz$.

A similar method can be used for integrating (1), where we have

$$Q(x,y,z) = P(y,x,z) \quad \text{and} \quad R(x,y,z) = R(y,x,z).$$

By treating x in (1) as a constant and integrating, we get (2) again: $U(x,y,z)=f(x)$, and by the symmetry we this time get

$$U(y,x,z) = f(y). \tag{5}$$

Solving (5) for z, substituting this value of z in (2), and giving y some suitable fixed value α, again gives us f.

One must again check that (1) is integrable before using this method.

Example: $yz(y-z)dx+xz(x-z)dy+xy(x+y)dz=0$. We have

$$\begin{vmatrix} yz(y-z) & xz(x-z) & xy(x+y) \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ yz(y-z) & xz(x-z) & xy(x+y) \end{vmatrix} = 0,$$

so the equation is integrable.

Treating x as a constant, we have $z(x-z)dy+y(x+y)dz=0$, and integration gives

$$\frac{(x+y)(x-z)}{yz} = f(x). \tag{6}$$

By symmetry we have $[(x+y)(y-z)]/xz=f(y)$. Putting $y=1$ and solving for z, we get

$$z = \frac{x+1}{(c_1+1)x+1} = \frac{x+1}{cx+1},$$

where $f(1)=c_1$ is an arbitrary constant and $c_1+1=c$.

Substituting $y=1$ in (6), we get

$$f(x) = \left\{x - \frac{x+1}{cx+1}\right\}(cx+1) = cx^2 - 1.$$

Thus the solution is

$$\frac{(x+y)(x-z)}{yz} = cx^2 - 1,$$

which can be rewritten in the more symmetric form $x+y-z=cxyz$.

A FINITE-VALUED FINITELY ADDITIVE UNBOUNDED MEASURE

D. P. GIESY, University of Southern California

0. INTRODUCTION. In trying to characterize the conjugate space (space of continuous linear functionals) of $\mathfrak{L}_\infty(X, \mathfrak{A}, \mu)$, the space of all μ-essentially bounded complex-valued $\mathfrak{A}$-measurable functions on X (see for example, "Real and Abstract Analysis" by Edwin Hewitt and Karl Stromberg, Springer-Verlag, New York, 1965, especially sections 20.14 and 20.27 ff), one quickly discovers

that each continuous linear functional L on $\mathcal{L}_\infty$ has the form

$$L(f) = \int_X f d\tau,$$

where τ is a complex-valued function on $\mathcal{A}$ satisfying:

(i) $\sup\{|\tau(A)| : A \in \mathcal{A}\} < \infty$ (τ is bounded on $\mathcal{A}$),

(ii) $\tau(A \cup B) = \tau(A) + \tau(B)$ if $A, B \in \mathcal{A}$ and $A \cap B = \phi$ (τ is finitely additive on $\mathcal{A}$), and

(iii) $\tau(A) = 0$ if $A \in \mathcal{A}$ and A is locally μ-null.

In my real analysis seminar, a student raised the question of whether (i) can be replaced by

(i′) $|\tau(A)| < \infty$ for each $A \in \mathcal{A}$ (τ is finite-valued on $\mathcal{A}$),

that is, whether (i′), (ii), and (iii) imply (i), as is the case for countably additive τ, where the even weaker "$|\tau(X)| < \infty$" can replace (i′). It is interesting that the answer to this question can be obtained (in the negative) by purely algebraic means.

I wish to thank Mr. Stephen Weinstein who brought this problem to my attention and Mr. Hal Forsey who suggested a simplification in my original proof.

1. LEMMA. *Let $\mathcal{B}$ and $\mathcal{C}$ be algebras of subsets of X, $\mathcal{B} \subset \mathcal{C}$, and let ρ be a finitely additive finite-valued measure on $\mathcal{B}$. Given $C \in \mathcal{C}$, $C \notin \mathcal{B}$, and a (finite) scalar, γ, there exists a finitely additive finite-valued measure τ on $\mathcal{C}$ such that $\tau | \mathcal{B} = \rho$ and $\tau(C) = \gamma$.*

Proof. If $\mathcal{L}$ is the linear space of all functions on X which are finite linear combinations of characteristic functions of sets in $\mathcal{B}$, and $\mathcal{M}$ is the linear space formed in the same way from $\mathcal{C}$, then $\mathcal{L} \subset \mathcal{M}$, and the characteristic function of C, χ_C, is in $\mathcal{M}$ but not $\mathcal{L}$. The fact that $\chi_C \notin \mathcal{L}$ follows from the fact that $\mathcal{B}$ is an algebra so that in the first place, each element of $\mathcal{L}$ can be represented using pairwise disjoint sets, and in the second place $C \notin \mathcal{B}$ so C is not a finite union of sets in $\mathcal{B}$. If f is defined on $\mathcal{L}$ by

$$f\left(\sum_{i=1}^{n} \beta_i \chi_{B_i}\right) = \sum_{i=1} \beta_i \rho(B_i)$$

then f is a well-defined linear functional on $\mathcal{L}$. The proof of this last statement is the same as the usual proof that defining $\int_X(\sum_{i=1}^n \alpha_i \chi_{A_i})d\mu$ to be $\sum_{i=1}^n \alpha_i \mu(A_i)$ yields a good definition of an integral on the set of simple functions supported by sets of finite measure and that this integral is a linear operator.

Then by extending a basis for $\mathcal{L}$ to a basis for $\mathcal{M}$ including χ_C in the extended basis, we construct a linear functional g on $\mathcal{M}$ which is an extension of f such that $g(\chi_C) = \gamma$. If we define τ by $\tau(S) = g(\chi_S)$ for each $S \in \mathcal{C}$, then it is easily verified that τ is the desired measure.

2. THEOREM. *Let $(X, \mathcal{A}, \mu)$ be a measure space such that there are infinitely*

many μ-essentially distinct sets of finite nonzero μ measure. Then there exists a function τ on $\mathcal{A}$ satisfying (i′), (ii), *and* (iii), *but not* (i).

Proof: Under the hypotheses of this theorem, it is easy to construct algebras $\mathcal{B}_0, \mathcal{B}_1, \mathcal{B}_2, \cdots$, and sets $B_1, B_2, \cdots$, such that

(a) $\mathcal{B}_0$ contains all locally μ-null sets in $\mathcal{A}$,

(b) $\mathcal{B}_0 \subset \mathcal{B}_1 \subset \mathcal{B}_2 \subset \cdots \subset \mathcal{A}$,

(c) for each $n \geqq 1$, $B_n \in \mathcal{B}_n$ but $B_n \notin \mathcal{B}_{n-1}$,

(d) for each $n \geqq 1$, $0 < \mu(B_n) < \infty$.

Define τ_0 on $\mathcal{B}_0$ to be the identically 0 function. Clearly τ_0 is a finitely additive finite-valued measure. Inductively, choose τ_n to be a Lemma 1 extension of τ_{n-1} from $\mathcal{B}_{n-1}$ to $\mathcal{B}_n$ such that $\tau_n(B_n) = n$. Finitely, let τ be a Lemma 1 extension of $\bigcup_{n=1}^{\infty} \tau_n$ from $\bigcup_{n=1}^{\infty} \mathcal{B}_n$ (clearly an algebra) to $\mathcal{A}$. Then τ is the desired measure.

3. Remarks. The restriction in Theorem 2 that X contain infinitely many μ-essentially distinct subsets of finite nonzero μ measure is easily seen to be necessary as well as sufficient to the conclusion of that theorem.

With notation as in the statement and proof of Theorem 2, if f is defined on the $\mathcal{A}$-simple functions on X by $f(s) = \int_X s\, d\tau$, then f is well-defined and linear (as in the proof of Lemma 1) but f is not continuous, since $n^{-1}\chi_{B_n} \to 0$ in $\mathcal{L}_\infty$ but $f(n^{-1}\chi_{B_n}) = 1$ for all n. Since the $\mathcal{A}$-simple functions are dense in $\mathcal{L}_\infty$, this gives an example of an unbounded linear functional on a dense subspace of a Banach space.

SOME RESULTS CONCERNING THE OCCURRENCE OF SPECIFIED PRIME FACTORS IN $\binom{m}{r}$

G. J. Simmons, University of New Mexico

It is well known that for any positive integers r and $n \geqq r$ the falling r-factorial of n

$$(n)_r = n(n-1) \cdots (n-r+1)$$

is always divisible by $r!$ Thus $p^h \mid (n)_r$ for each prime $p \leqq r$, where p^h with $h \geqq 0$ is the highest power of p dividing $r!$, written $p^h \| r!$. This note shows conversely that there are infinitely many n such that $p^l \mid r!$ for each prime $p \leqq r$, where $p^l \| (n)_r$. The even stronger result is true that if N and r are arbitrary integers, there are infinitely many m such that

$$\left(\binom{m}{r}, N\right) = 1. \tag{1}$$

It is known that the binomial coefficient $\binom{m}{r}$ is not divisible by a fixed prime p if and only if when r and $m-r$ are represented in p-ary notation, forming their sum does not generate a carry in any digit position. This result was first given by E. Kummer [1] and later extended to multinomial coefficients by L. E.

Dickson [2] in his dissertation at the University of Chicago and by J. W. L. Glaisher [3, 4]. If r and $m-r$ satisfied the Kummer conditions for each prime factor of N, then the conclusion given by (1) would be proved. A direct proof of the existence of m with these properties by this method is possible, but obscure.

LEMMA. *Given any integers $n \geqq r > 0$ and an arbitrary prime p, let p^h be determined by $p^h \| (n)_r$. Then $p^h \| (n+kp^{h+1})_r$ for each nonnegative integer k.*

Proof.

$$(n + kp^{h+1})_r = \prod_{j=0}^{r-1} \{(n-j) + kp^{h+1}\},$$

hence

$$(n + kp^{h+1})_r \equiv \prod_{j=0}^{r-1} (n-j) \equiv (n)_r \pmod{p^{h+1}}.$$

THEOREM. *Given any positive integers r and N, there exist infinitely many integers $m \geqq r$ such that*

$$\left(\binom{m}{r}, N\right) = 1.$$

Proof. Let $m = r + kN(r!)^2$ where k is an arbitrary nonnegative integer, and let p be any prime factor of N. We prove that

$$p \nmid \binom{m}{r}.$$

First, if $p \nmid r!$, i.e., $p \nmid (r)_r$, then by the lemma with $n = r$ and $h = 0$, we see that $p \nmid (m)_r$ and hence

$$p \nmid \binom{m}{r}.$$

Second, if $p \mid r!$, let h be determined by $p^h \| r!$, so that $p^{h+1} \mid (r!)^2$ since $h \geqq 1$. Then by the lemma with $n = r$ we see that $p^h \| (m)_r$, and hence

$$p \nmid \binom{m}{r}.$$

This work was supported by the United States Atomic Energy Commission.

References

1. E. E. Kummer, Über die Ergänzungssätze zu den allgemeinen Reciprocitätsgesetzen, J. Reine Angew. Math., 44 (1852) 93–146.

2. L. E. Dickson, Analytic representation of substitutions, Ann. of Math., II (1896–97) 75–76.

3. J. W. L. Glaisher, On the residue of a binomial-theorem coefficient with respect to a prime modulus, Quart. J. Pure and Appl. Math., 30 (1899) 150–156, 349–366.

4. ———, Theorems on the residues of multinomial coefficients with respect to a prime modulus, Quart. J. Pure and Appl. Math., 33 (1902) 378–384.

MATHEMATICAL EDUCATION

EDITED BY J. G. HARVEY AND M. W. POWNALL

Material for this Department should be sent to either of the editors: J. G. Harvey, Department of Mathematics, University of Wisconsin, Madison, WI 53706; M. W. Pownall, Department of Mathematics, Colgate University, Hamilton, NY 13346.

MEETING REPORT: CONFERENCE ON SCIENCE IN THE TWO-YEAR COLLEGE

Thirty-two representatives of seventeen national scientific and educational organizations* convened in Washington on June 18–19, 1969 for a *Conference on Science in the Two-Year College* that was supported by the National Science Foundation and sponsored by the Commission on Undergraduate Education in the Biological Sciences (CUEBS). This meeting followed two earlier conferences called by the college science commissions, and provided an opportunity for a common focus of the independent two-year college efforts of all the organizations represented.

Following an address by Dale Tillery of the Center for Research and Development in Higher Education, University of California at Berkeley, the conferees addressed themselves to the development of recommendations on two-year college science with particular reference to matters of administration, curriculum, and personnel. The specific consensus of recommendations, several of which were reached without unanimity, have been distributed to the conferees for transmittal to their respective organizations for endorsement and implementation. By late spring 1970, the endorsed guidelines, with a supporting rationale, will be disseminated widely to regional accrediting associations, professional organizations, two-year college administrators, teachers, and so on, for further implementation and to serve as a basis for continuing dialogue among groups concerned with science in two-year colleges.

* American Association for the Advancement of Science, American Association of Junior Colleges, Advisory Council on College Chemistry, American Council on Education, American Chemical Society, American Geological Institute, American Institute of Biological Sciences, American Institute of Physics, Commission on Education of the National Academy of Engineering, Commission on Undergraduate Education in the Biological Sciences, Council on Education in the Geological Sciences, Committee on the Undergraduate Program in Mathematics, Mathematical Association of America, National Faculty Association of Community and Junior Colleges (representing the National Education Association), National Science Foundation.

PROBLEMS AND SOLUTIONS

All problems (both elementary and advanced) proposed for inclusion in this Department should be sent to E. P. Starke, 1000 Kensington Ave., Plainfield, N.J. 07060. Proposers of problems are urged to enclose any solutions or information that will assist the editors. Ordinarily, problems in well-known textbooks and results in generally accessible sources are not appropriate for this Department. No solutions (except those accompanying proposals) should be sent to Professor Starke.

ELEMENTARY PROBLEMS

Solutions of Elementary Problems should be sent to Problems Group, Mathematics Department, University of Maine, Orono, ME 04473. To facilitate their consideration, solutions of Elementary Problems in this issue should be typed (with double spacing) and should be mailed before August 31, 1970. Contributors (in the United States) who desire acknowledgement of receipt of their solutions are asked to enclose self-addressed stamped postcards.

E 2235. *Proposed by Gideon Nettler, Montclair (N.J.) State College*

Prove that

$$\tau^2(d) = \sum_{c|d} \sum_{b|c} \sum_{a|b} \mu^2(a),$$

where $\mu(n)$ is the Möbius function and $\tau(n)$ is the number of divisors of n.

E 2236. *Proposed by M. S. Klamkin, Ford Scientific Laboratory and D. J. Newman, Yeshiva University*

Show that if the integral of the reciprocal of a nonconstant polynomial is a rational function, then the polynomial must be of the form $(ax+b)^n$.

E 2237. *Proposed by Arthur Marshall, Madison, Wisconsin*

Let n be an odd, composite, square-free natural number, such that no prime factor divides the Euler totient of any other prime factor. Prove that $k\phi(n) = n-1$ has no solutions for any natural number k. (This problem was posed in D. H. Lehmer, *On Euler's totient function*, Bull. A.M.S., 38(1932) 745–751.)

E 2238. *Proposed by Emanuel Vegh, U. S. Naval Research Laboratory*

Let p be a prime and $t_1, t_2, \cdots, t_n$ a reduced residue system mod $p-1$. If g is a primitive root of p, then it is well known that the integers $g^{t_1}, g^{t_2}, \cdots, g^{t_n}$ are distinct mod p. (In fact, these integers are the primitive roots of p.) Is

there an integer h, *not* a primitive root of p, such that $h^{t_1}, h^{t_2}, \cdots, h^{t_n}$ are distinct mod p?

E 2239. *Proposed by G. Sabbagh, Paris, France*

Let R be a binary relation on a nonvoid set E such that:

(1) For no $x \in E$ is xRx true.

(2) For each pair (x, y) of distinct elements of E one and only one of the following relations holds: xRy, yRx.

(3) R is dense, which means: If xRy then there is a $z \in E$ such that xRz and zRy. Must E be infinite?

E 2240. *Proposed by Yeong-shyeong Tsai, Taiwan Provincial Chung-Hsing University*

Let n be an arbitrary positive integer, and let A be the set $\{1, 2, 3, \cdots, n\}$. Take n vertical line segments between two horizontal lines, indexing the top points and the bottom points by the numbers $1, 2, \cdots, n$. Now draw as many horizontal line segments as you please between adjacent vertical lines, with the restriction that all endpoints of these segments are to be distinct. Now, starting at any upper point i, proceed downward until the first horizontal segment is reached, then go across that segment and again proceed downward; continue in this way, passing over each horizontal segment as it is met until a bottom point, say k_i, is reached.

(1) Prove that the mapping $i \rightarrow k_i$ $(i = 1, 2, \cdots, n)$ is a one-to-one mapping of A onto A.

(2) Prove that in completing the mapping $i \rightarrow k_i$ for all i as described above, every part of a vertical line is traced exactly once, and every horizontal segment is traced twice, once in each direction.

E 2241. *Proposed by Linda Pleska, Bowling Green State University*

Prove or disprove: If $\lim_{x \to x_0} f(x)$ exists for each $x_0 \in [a, b]$, then the Riemann integral $\int_a^b f(x)dx$ exists.

SOLUTIONS OF ELEMENTARY PROBLEMS

The following solutions were sent in response to the list of unsolved problems published in Vol. 76, No. 6, p. 711. The editors will welcome further contributions.

Another Inequality for the Altitudes and Exradii of a Triangle

E 1847 [1966, 81]. *Proposed by J. I. Nassar, Muhlenberg College*

Let h_i, r_i $(i = 1, 2, 3)$ be the altitudes and exradii respectively of a triangle, and let $\alpha_1, \alpha_2, \alpha_3$ be any nonnegative numbers. Prove that

$$\sum h_i^{\alpha_1} h_j^{\alpha_2} h_k^{\alpha_3} \leqq \sum r_i^{\alpha_1} r_j^{\alpha_2} r_k^{\alpha_3},$$

where the summation is taken over all permutations of 1, 2, 3. The equality

holds if and only if the triangle is equilateral. Here x^α is taken to be the positive number x^α for $x>0$, $\alpha\geqq 0$. This is an extension of E 1675 [1965, 187].

Solution by Anders Bager, Hjørring, Denmark. It is clear that we must avoid the case $\alpha_1=\alpha_2=\alpha_3=0$ if we wish to preserve the stated condition of equality.

In a more usual notation we prove the following slight generalization. Let α, β, γ be nonnegative numbers, not all zero. Then

$$(1)\qquad h_a^\alpha h_b^\beta h_c^\gamma + h_b^\alpha h_c^\beta h_a^\gamma + h_c^\alpha h_a^\beta h_b^\gamma \leqq r_a^\alpha r_b^\beta r_c^\gamma + r_b^\alpha r_c^\beta r_a^\gamma + r_c^\alpha r_a^\beta r_b^\gamma$$

with equality if and only if $a=b=c$. By circling the other way round the triangle we get another inequality of this sort, and by adding the two we get the stated inequality with the condition of equality preserved.

We set $s_a=-a+b+c$, $s_b=a-b+c$, $s_c=a+b-c$ and get

$$(2)\qquad \tfrac{1}{2}(s_a+s_b)=c,\qquad \tfrac{1}{2}(s_b+s_c)=a,\qquad \tfrac{1}{2}(s_c+s_a)=b.$$

As $ah_a=bh_b=ch_c=r_as_a=r_bs_b=r_cs_c$ (double area of the triangle) (1) may be rewritten as

$$(3)\quad a^{-\alpha}b^{-\beta}c^{-\gamma}+b^{-\alpha}c^{-\beta}a^{-\gamma}+c^{-\alpha}a^{-\beta}b^{-\gamma}\leqq s_a^{-\alpha}s_b^{-\beta}s_c^{-\gamma}+s_b^{-\alpha}s_c^{-\beta}s_a^{-\gamma}+s_c^{-\alpha}s_a^{-\beta}s_b^{-\gamma}.$$

Clearly we need only consider the following cases:

(i) $\alpha=\beta=\gamma>0$. By (2) and the theorem of the arithmetic and the geometric mean we get $(s_as_b)^{1/2}\leqq c$ (equality if and only if $a=b$), and two analogous inequalities. We multiply these and get

$$s_as_bs_c\leqq abc,\qquad s_a^{-\alpha}s_b^{-\alpha}s_c^{-\alpha}\geqq a^{-\alpha}b^{-\alpha}c^{-\alpha},$$

which is (3) in this case (with equality if and only if $a=b=c$).

(ii) $\alpha>0$, $\beta=\gamma=0$. As the function $f(x)=x^{-\alpha}(x>0)$ is convex, we get $f(s_a)+f(s_b)\geqq 2f(c)$ (equality if and only if $a=b$) and two analogous inequalities. By adding the three inequalities and dividing by 2 we get

$$f(s_a)+f(s_b)+f(s_c)\geqq f(a)+f(b)+f(c)$$

(equality if and only if $a=b=c$), which is (3) in this case.

(iii) $\alpha>0,\beta>0,\gamma=0$. Here we use Theorem 99, p. 80, in Hardy-Littlewood-Pólya, *Inequalities* (Cambridge, 1964). Consider the function $\Phi(x,y)=x^{-\alpha}y^{-\beta}$ defined in the first quadrant $(x>0,y>0)$. We get

$$\Phi_{xx}(x,y)=\alpha(\alpha+1)x^{-\alpha-2}y^{-\beta}>0$$

for all x, y, and

$$(\Phi_{xx}\Phi_{yy}-\Phi_{xy}^2)(x,y)=\alpha\beta(\alpha+\beta+1)x^{-2\alpha-2}y^{-2\beta-2}>0$$

for all x, y. Hence

$$\Phi(s_a,s_b)+\Phi(s_b,s_c)\geqq 2\Phi(c,a)$$

(equality if and only if $a=b=c$) and two analogous inequalities. By adding the three inequalities and dividing by 2 we get

$$\Phi(s_a, s_b) + \Phi(s_b, s_c) + \Phi(s_c, s_a) \geqq \Phi(a, b) + \Phi(b, c) + \Phi(c, a)$$

(equality if and only if $a=b=c$), which is (3) in this case.

(iv) $\alpha>\gamma>0$, $\beta\geqq\gamma$. By earlier cases we get

$$a^{-\gamma}b^{-\gamma}c^{-\gamma} \leqq s_a^{-\gamma}s_b^{-\gamma}s_c^{-\gamma}, \tag{4}$$

$$\begin{aligned} &a^{-(\alpha-\gamma)}b^{-(\beta-\gamma)} + b^{-(\alpha-\gamma)}c^{-(\beta-\gamma)} + c^{-(\alpha-\gamma)}a^{-(\beta-\gamma)} \\ &\qquad \leqq s_a^{-(\alpha-\gamma)}s_b^{-(\beta-\gamma)} + s_b^{-(\alpha-\gamma)}s_c^{-(\beta-\gamma)} + s_c^{-(\alpha-\gamma)}s_a^{-(\beta-\gamma)}. \end{aligned} \tag{5}$$

We multiply (4) and (5) and get (3), with the usual condition for equality.

Note. In the original statement of the problem the sign of inequality was unfortunately reversed.

Nesting Habits of the Laddered Parenthesis

E 1903 [1966, 666]. *Proposed by George Eldredge, El Cerrito, California*

Let an n-ladder of twos, L_n, be defined as follows:

$$L_n = 2^{2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}}$$

where there are n twos. Let N_n be the number of distinct integers that can be obtained from L_n by the appropriate insertion of a set of unambiguous nested parentheses. For example, $N_3=1$, $N_4=2$. Find N_n.

Solution by Michael Goldberg, Washington, D. C. Since

$$(2^2)^2 = 2^{(2^2)} = 16,$$

the value is independent of the placing of the parentheses, and $N_3=1$. Also

$$\left(2^{(2^2)}\right)^2 = 16^2 = 256, \qquad \text{while } 2^{(2^{(2^2)})} = 2^{16}.$$

Hence, $N_4=2$. Expressions like

$$(2^{2^2)^{(2^{2^2)}}}$$

are excluded since the parentheses are not nested. Hence each new two must be added at only the top or bottom of the ladder of lower order.

For $n=5$, we have the values

$$(256)^2 = (2^8)^2 = 2^{16}; \qquad (2^{16})^2 = 2^{32}; \qquad 2^{256}; \qquad 2^{(2^{16})}$$

and $N_5=4$.

For larger values of n, the values obtained by adding the two at the top of the ladder fall very much below the values obtained by adding the two at the bottom of the ladder. Hence, $N_{n+1}=2N_n$ or $N_n=2^{(n-3)}$ for $n=3, 4, 5, \cdots$.

Note. This solution, originally submitted in 1966 was among a number which were inadvertently lost.

A Sequence Problem with a Unique Solution

E 1905 [1966, 774]. *Proposed by R. L. Graham and L. A. Shepp, Bell Telephone Laboratories*

Let $x_1=x$, $x_{n+1}=x^{x_n}$ for $n=1, 2, \cdots$. If $a>e^{1/e}$, prove there is a γ for which $\lim_{n\to\infty}\gamma_n/a_{n+1}$ exists and is a positive number. Is γ unique?

Solution by J. H. Lindsey, II, the RAND Corporation. Fix $a>e^{1/e}$. One can easily show by induction that for $\gamma>1$, γ_n is monotonic increasing. If $\lim_{n\to\infty} a_n$ exists, say $\alpha=\lim_{n\to\infty} a_n$, then $\lim_{n\to\infty} a_n=\alpha=\lim_{n\to\infty} a_{n+1}=\lim_{n\to\infty} a^{(a_n)}=a^\alpha$ and $a=\alpha^{(1/\alpha)}\leqq e^{(1/e)}$, a contradiction. Therefore, the sequence a_n is monotone increasing to infinity.

Now $a_n<a_{n+1}$ and $(a_{n+1})_n>(a_{n+1})_1=a_{n+1}$. As γ_n is a continuous function of γ for $\gamma>1$, by the intermediate value theorem, we may find $f(n)$ with $a<f(n)<a_{n+1}$ and $(f(n))_n=a_{n+1}$. If $b>c>1$ then one can show by induction that $b_n>c_n$. Therefore, $f(n)$ is uniquely determined by a and n.

For any $\gamma>1$,

$$\frac{\gamma_{n+1}}{a_{n+2}}=\frac{\gamma^{(\gamma_n)}}{a^{(a_{n+1})}}=\frac{\gamma^{(\gamma_n)}}{(\gamma^{(\log_\gamma a)})^{a_{n+1}}}=\gamma^{\phi a_{n+1}},$$

where $\phi=(\gamma_n/a_{n+1})-\log_\gamma a$. Therefore,

$$\frac{\gamma_n}{a_{n+1}}=\log_\gamma a+\frac{1}{a_{n+1}}\log_\gamma\left(\frac{\gamma_{n+1}}{a_{n+2}}\right).$$

In particular,

$$\frac{(f(n))_{n-1}}{a_n}=\log_{f(n)} a+\frac{1}{a_n}\log_{f(n)}\left(\frac{(f(n))_n}{a_{n+1}}\right)=\log_{f(n)} a<1.$$

As $(f(n-1))_{n-1}/a_n=1$, we have $f(n-1)>f(n)$.

Since the $f(n)$ are decreasing and bounded below by a, we may write $x=\lim_{n\to\infty} f(n)$ for some $x\geqq a$.

We shall show that $\lim_{n\to\infty} x_n/a_{n+1}=\log_x a$. First,

$$x_{n-1}/a_n<(f(n))_{n-1}/a_n=\log_{f(n)} a<\log_x a.$$

Let $\log_x a>\epsilon>0$. As

$$\lim_{m\to\infty}(\log_{f(m)} a)=\log_x a \quad\text{and}\quad \lim_{n\to\infty} a_n=\infty,$$

there exist M and N so that $m>M$ and $n>N$ together imply

$$[\log_x(a) - \epsilon] - \left[\log_{f(m)} a + \frac{1}{a_{n+1}} \log_{f(m)}(\log_x (a) - \epsilon)\right] < 0.$$

For $m>n$, $m>M$, and $n>N$, we shall show successively that $(f(m))_i/a_{i+1} >\log_x (a)-\epsilon$ for $i=m, m-1, \cdots, n$. This is true for $m=i$, since $(f(m))_m/a_{m+1} =1>\log_x (a)-\epsilon$. If the statement is true for $i+1$ with $i\geqq n$, then

$$\begin{aligned}\frac{(f(m))_i}{a_{i+1}} &= \log_{f(m)} a + \frac{1}{a_{i+1}} \log_{f(m)}\left(\frac{(f(m))_{i+1}}{a_{i+2}}\right) \\ &> \log_{f(m)} a + \frac{1}{a_{i+1}} [\log_{f(m)}(\log_x (a) - \epsilon)] \\ &> \log_x (a) - \epsilon.\end{aligned}$$

Therefore, for $n>N$, $m>M$, and $m>n$, $(f(m))_n/a_{n+1}>\log_x (a)-\epsilon$. Taking the limit of this inequality as $m\to\infty$, we have, for $n\geqq N$, $x_n/a_{n+1}\geqq\log_x (a)-\epsilon$. Therefore

$$\lim_{n\to\infty} \frac{x_n}{a_{n+1}} = \log_x a.$$

Suppose $\lim_{n\to\infty} (\gamma_n/a_{n+1})=c(\gamma)$, a positive number. Then we may let n approach infinity in the equality

$$\frac{\gamma_n}{a_{n+1}} = \log_\gamma a + \frac{1}{a_{n+1}} \log_\gamma \frac{\gamma_{n+1}}{a_{n+2}}$$

and conclude $c(\gamma)=\log_\gamma a$. Suppose that there are two solutions x and y for γ with $x<y$. Then $1<x$ and $c(x)=\log_x a>\log_y a=c(y)$. However,

$$c(x) = \lim_{n\to\infty} \frac{x_n}{a_{n+1}} \leqq \lim_{n\to\infty} \frac{y_n}{a_{n+1}} = c(y),$$

a contradiction from which we can conclude the uniqueness of γ.

An Unordered Pair Undefined

E 1917 [1966, 891]. *Proposed by K. O. May, University of Toronto*

The ordered pair is customarily defined in terms of the unordered pair by $(a, b)=\{\{a, b\}, \{a\}\}$. Prove that it is impossible to reverse this by defining unordered pair in terms of ordered pair.

Solution by D. W. Hadwin, Augustana College, South Dakota. There are two parts of this problem which need clarification. First I will assume that when the proposer says "a redefinition of the unordered pair $\{a, b\}$," he means an expression involving a and b which is unaltered by interchanging a and b. The most ambiguous point in this problem is the phrase "in terms of." In a sense $\{(a, b), (b, a)\}$ is an expression written in terms of the ordered pairs (a, b) and

(b, a) which is unaltered by interchanging a and b. In view of this example I am led to believe that the proposer's intentions will be expressed in the reformulation of the problem given below.

Define a sequence of sets $\{S_n\}$ inductively by:

$$S_0 = \{a, b\}, \qquad S_{n+1} = S_n \cup \{(u, v) \mid u, v \in S_n\}.$$

Then S_n is the set of all possible expressions which can be formed from a and b using at most n ordered pairs. Let $S = \bigcup_{n=1}^{\infty} S_n$ and define a mapping $x \to \bar{x}$ from S to S inductively as follows: $\bar{a} = b$, $\bar{b} = a$, and if $(u, v) \in S_n$ then $\overline{(u, v)} = (\bar{u}, \bar{v})$. Note that $\bar{x}$ is merely the expression obtained from x by interchanging a and b.

The original problem can now be restated: there is no $x \in S$ for which $x = \bar{x}$.

Proof. By way of contradiction assume that there is an $x \in S$ such that $x = \bar{x}$. Let n be the smallest nonnegative integer for which there is an $x \in S_n$ such that $x = \bar{x}$ and let $x_0 \in S_n$ be such an x. Clearly $n \neq 0$. Since $x_0 \in S_n$ and $x_0 \notin S_{n-1}$, there exist $u, v \in S_{n-1}$ such that $x_0 = (u, v)$. Thus $x_0 = (u, v) = \bar{x}_0 = (\bar{u}, \bar{v})$. Hence $u = \bar{u}$, $v = \bar{v}$, contradicting the minimality of n.

Also solved by Gerald Edgar, S. J. Garland, and by J. H. Lindsey, II.

Edgar shows that the result holds if we replace "ordered pair" by any function f such that if $f(x, y) = f(u, v)$, then $x = u$.

Derivatives of Periodic Functions

E 1959 [1967, 198]. *Proposed by R. A. Struble, North Carolina State University*

If $f(x, y) \in C^1$ where f is periodic in x with least (positive) period $P(y)$, then the partial derivative $f_y(x, y)$ is periodic if and only if $P'(y) = 0$.

I. *Comment by D. A. Hejhal, University of Chicago.* The "only if" part is false without some additional hypothesis. Consider the following counterexample. Let $f(x, y) = y^2 \sin x + e^y \sin 2x$. Then $P(0) = \pi$ but $P(y) = 2\pi$ for all sufficiently small $y \neq 0$. Thus $P(y)$ is discontinuous at $y = 0$, whence $P'(0)$ does not exist.

II. *Solution of the problem amended by the assumption that $P(y)$ is continuous.*

LEMMA 1: If $P(y)$ is continuous, then it is differentiable.

Proof by the proposer. For all x, y, Δy we have

$$\begin{aligned}(x + P(y + \Delta y), y) &- f(x + P(y), y) \\ &= f(x + P(y + \Delta y), y) - f(x, y) \\ &= f(x + P(y + \Delta y), y) - f(x + P(y + \Delta y), y + \Delta y) + f(x, y + \Delta y) - f(x, y).\end{aligned}$$

Thus, using the mean value theorem, we have further

$$\begin{aligned}f_x(x + P(y) + \xi, y)[P(y + \Delta y) - P(y)]& \\ = - f_y(x + P(y + \Delta y), y + \eta)\Delta y + f_y(x, y + \zeta)\Delta y&\end{aligned} \tag{1}$$

for suitable ξ, η and ζ satisfying $|\xi| \leqq |P(y+\Delta y)-P(y)|$, with $|\eta| \leqq |\Delta y|$ and $|\zeta| \leqq |\Delta y|$. Note that

$$(2) \qquad |\xi| \to 0 \qquad \text{as } |\Delta y| \to 0 \text{ (uniformly in } x)$$

since $P(y)$ is continuous. Now we consider y fixed and choose x_0 which satisfies $f_x(x_0, y) \neq 0$. This is possible since $f(x, y)$ has a *least positive* period in x. Moreover, since $f_x(x, y)$ is continuous, there exists a $\delta > 0$ such that

$$(3) \qquad |x_0 - z| < \delta \qquad \text{implies } f_x(z, y) \neq 0.$$

Because of (2), there exists a $\delta_1 > 0$ such that

$$(4) \qquad |\Delta y| < \delta_1 \qquad \text{implies } |\xi| < \delta \quad \text{for all } x.$$

Hence for each Δy satisfying $|\Delta y| < \delta_1$, there exists a number x_1 satisfying

$$(5) \qquad |x_1 + P(y) + \xi - x_0| < \delta,$$

where $\xi = \xi(x_1, \Delta y)$ satisfies (1) with $x = x_1$. Moreover, because of (2), we may impose the condition $x_1 \to x_0 - P(y)$ as $|\Delta y| \to 0$. Hence, using (1), (3) and (5), for $0 < |\Delta y| < \delta_1$, we may write

$$\frac{P(y + \Delta y) - P(y)}{\Delta y} = \frac{-f_y(x_1 + P(y + \Delta y), y + \eta) + f_y(x_1, y + \zeta)}{f_x(x_1 + P(y) + \xi, y)},$$

where the right hand member possesses a limit as $|\Delta y| \to 0$. Thus $P(y)$ is differentiable.

LEMMA 2: The assertion of the original problem is true if $P(y)$ is differentiable.

Proof by Paul Nelson, Jr., University of New Mexico. We may differentiate the identity

$$f(x + P(y), y) = f(x, y)$$

with respect to y, obtaining

$$f_x(x + P(y), y)P'(y) = f_y(x, y) - f_y(x + P(y), y).$$

The result now follows from the observation that $f_x(x+P(y), y)$ is not identically zero in x.

Some Nonregular Polyhedra

E 2141 [1969, 82, 1067]. *Proposed by H. S. Hahn, West Georgia College*

(1) Find convex nonregular equilateral (equal-edged) polyhedra with their n $(4 < n < 16)$ vertices on a sphere.

(2) Prove or disprove that there is no convex nonregular equilateral polyhedron with an odd number n of vertices all on a sphere, except for $n = 5$, 9, 11, 15, and 55.

Comments by Michael Goldberg, Washington, D. C. The editorial note with

the original solution indicated that no proof or disproof of part (2) was attempted. Johnson's paper in the *Canadian Journal of Mathematics* 18(1966), 169–200, tabulated all the convex polyhedra with regular faces, but no rigorous proof was given. Since then, however, a rigorous and laborious proof was published in the U.S.S.R. It has been translated from the Russian and published as *Convex Polyhedra with Regular Faces*, V. A. Zalgaller, Consultants Bureau, New York, 1969.

A Nonsingular Matrix?

E 2184 [1969, 825]. *Proposed by J. C. Nichols, Monmouth (Illinois) College and (independently) by J. A. Huckaba, University of Missouri*

Let a_{ij} be nonnegative real numbers such that (a) $a_{ii}=1$, (b) $a_{ij}a_{ji}<1$, and (c) $a_{ij}a_{jk}\leqq a_{ik}$ for all i, j with $1\leqq i, j\leqq n$ and $i\neq j$.

Show that the matrix $[a_{ij}]$ has a positive determinant for $n=3$. This is obviously true for $n=1, 2$; can you establish this result for $n>3$? At least show that $[a_{ij}]$ is non-singular for all n.

I. *Solution by E. F. Schmeichel, College of Wooster.* Let $D=\det[a_{ij}]$. Multiply the first row of D by $a_{21}a_{31}$, the second row by a_{31}, and the third by a_{21}. Now remove the factor $a_{21}a_{31}$ from the first column, subtract the second row from each of the others, and remove the factor a_{31} from the resulting first row. This yields

$$D=\frac{1}{a_{21}}\begin{vmatrix} a_{32}a_{21}-a_{31} & a_{21}-a_{23}a_{31} \\ a_{21}a_{12}-1 & a_{21}a_{13}-a_{23}\end{vmatrix}\equiv\frac{1}{a_{21}}\begin{vmatrix} b_{11} & b_{12} \\ b_{21} & b_{22}\end{vmatrix}.$$

Note that $b_{11}\leqq 0$, $b_{22}\leqq 0$, $b_{12}\geqq 0$, $b_{21}<0$. If $b_{12}>0$, we have $D=(b_{11}b_{22}-b_{12}b_{21})/a_{21}\geqq -b_{12}b_{21}/a_{21}>0$. If $b_{12}=0$, then $a_{21}=a_{23}a_{31}$, and so $b_{11}=(a_{32}a_{23})a_{31}-a_{31}<0$, since $a_{32}a_{23}<1$. Similarly $b_{22}<0$. Therefore $D=(b_{11}b_{22}-0\cdot b_{21})/a_{21}>0$, as required.

Noting that

$$\begin{bmatrix} 1 & 2(1-1/n) & 1 \\ \frac{1}{4} & 1 & \frac{1}{2} \\ \frac{1}{2} & 2(1-1/n) & 1\end{bmatrix},$$

where n is a positive integer, fulfills the given conditions, and has determinant $1/n^2$, we see that 0 is the best possible lower bound.

II. *Solution by L. J. Lange, University of Missouri.* The matrix $[a_{ij}]$ may be singular for $n=6$.

Let $A=[a_{ij}]$ be an $n\times n$ circulant of real numbers where $a_{ij}=a_{j-i}$ and, for $j<i$, $a_{j-i}=a_{n+j-i}$. If ω_t $(t=1, 2, \cdots, n)$ are the n nth roots of unity, it is well known that the characteristic roots of A are

$$\alpha_t=\sum_{k=0}^{n-1}a_k\omega_t^k.$$

If, in particular, $n=6$, $\omega_t=-1$, $a_0=1$, $a_1=2/3$, $a_2=1/2$, $a_3=2/3$, $a_4=1/2$, $a_5=2/3$, one easily sees that the conditions of the problem are met and $\alpha_t=0$.

Also solved by J. C. Nichols.

Editorial Note. Modification of Lange's method will yield singular matrices for all even n greater than 6. For $n=2k$, let $a_0=1$, $a_{2i+1}=(k-1)/k$ $(i=0, 1, \cdots, k-1)$, and $a_{2i}=(k-2)/(k-1)$ $(i=1, 2, \cdots, k-1)$.

Minimal Curve for Fixed Area

E 2185 [1969, 825]. *Proposed by Michael Goldberg, Washington, D. C.*

Given a convex quadrilateral. Find the shortest curve which divides it into two equal areas.

I. *Solution by C. S. Ogilvy, Hamilton College.* It is well known that the minimal curve which, together with two sides of an angle, encloses a fixed area, forms a circular sector. (Courant and Robbins, *What is Mathematics?*, p. 502, No. 97.) This implies that the solution curve is that circular arc perpendicular to two sides of the quadrilateral that halves the area. Let A be such an arc perpendicular to two adjacent sides and B be such an arc perpendicular to two opposite sides. Depending on shape, some quadrilaterals will have solutions of type A, some of type B. There may be transitional cases with equal solution arcs, one A and one B. (The square has two equal B solutions, both straight lines.) To characterize the shapes leading to each solution may necessitate numerical calculations.

II. *Solution by the proposer.* The curve may be considered as a restraining member under tension produced by internal fluid pressure in the restricted area. The ends of the curve are free to slide along the sides. Hence, the curve must be normal to two sides of the quadrilateral. Furthermore, since the fluid pressure is uniform, the curve must take the form of a circular arc. With the foregoing conditions in mind, the various possibilities must be investigated separately for each shape of quadrilateral. The method may be generalized to other polygons. Also, it may be extended, with further complications, to nonconvex polygons.

Also solved by M. T. Bird & V. E. Hoggatt, Jr., and Bill Sands. One unsigned solution was received.

Self-inverse Functions

E 2186 [1969, 825]. *Proposed by R. G. Kuller, Northern Illinois University*

Let g be a continuous real valued function defined on a real interval J. Assume that g is not the identity function I, but that for some integer $m \geqq 2$, $g \circ g \circ \cdots \circ g = I$, where there are m g's and the symbol $\circ$ denotes composition. Show that $g \circ g = I$.

Solution by Douglas Lind, Stanford University. Suppose $g: J \to J$ is continuous with $g^m = g \circ \cdots \circ g = I$. Then $g(x) = g(y)$ implies $x = g^m(x) = g^m(y) = y$, so g is injective and hence strictly monotone since g is continuous. If g is increasing,

then assuming $g(x)>x$ leads to $x=g^m(x)>g^{m-1}(x)> \cdots >g(x)>x$, and a similar contradiction results upon assuming $g(x)<x$. Hence $g(x)=x$ for all x in J, so $g=I$. If g is decreasing, then $x<y$ implies $g(x)>g(y)$, so $g^2(x)<g^2(y)$, showing g^2 is increasing. Since $(g^2)^m=(g^m)^2=I^2=I$, $g^2=I$ by the first case.

Also solved by D. M. Bloom, John Bryant & Robert Gilmer, Orin Chein, E. H. Davis, E. P. Del Norte, N. G. Fine, M. H. Fish, Kenneth Fogarty, W. F. Fox, Gilles Gauthier, M. G. Greening (Australia), C. V. Heuer & G. A. Heuer, A. E. Hoffman, Eleanor G. Jones, J. F. Leetch, Gesing Leung (Hong Kong), Andrej Mąkowski (Poland), Dan Marcus, J. V. Michalowicz, Simeon Reich (Israel), A. A. Sardinas, E. F. Schmeichel, D. E. Smith, R. A. Struble, A. C. Williams, Mark Yu, and the proposer.

Gauthier shows that the assertion holds for any Darboux function on set E to E. Mąkowski observes that the result follows at once from Theorems 15.2 and 15.3 of Kuczma, *Functional Equations of a Single Variable* (Monografie Matematyczne, vol. 46).

ADVANCED PROBLEMS

All solutions of Advanced Problems should be sent to Joshua Barlaz, Rutgers—The State University, New Brunswick, N. J. 08903. Solutions of Advanced Problems in this issue should be typed (with double spacing) on separate, signed sheets and should be mailed before August 31, 1970. Contributors (in the United States) who desire acknowledgement of receipt of their solutions are asked to enclose self-addressed stamped postcards.

An asterisk (*) means neither the proposer nor the editors supplied a solution.

5734.* *Proposed by G. G. Lorentz, University of Texas*

Let E_n be the degree of uniform approximation of x^{n+1} on the interval $[0, 1]$ by polynomials

$$(\dagger) \qquad P_n(x) = \sum_{k=0}^{n} a_k x^k (1-x)^{n-k}, \qquad a_k \geqq 0$$

with positive coefficients in $x(1-x)$. In other words, let E_n be the minimum of $\max_{0\leq x\leq 1}|x^{n+1}-P_n(x)|$ for all polynomials of the form (†). Prove that $y=\lim_{n\to\infty} nE_n$ exists, and find y.

5735.* *Proposed by Paul Erdös and T. Motzkin, University College of Swansea, Wales*

Denote by $F(n)$ the number of pairs $1\leqq a\leqq b\leqq n$ for which a and b have the same prime factors. Prove that $\lim_{n\to\infty} F(n)/n$ exists and is finite.

5736. *Proposed by M. S. Klamkin, Ford Scientific Laboratory*

Solve the nonlinear difference equation of rth order

$$D_n = a_1 D_{n-1}^{m+1} + a_2 D_{n-1}^{m} D_{n-2}^{m+1} + \cdots + a_r D_{n-1}^{m} D_{n-2}^{m} + \cdots + D_{n-r+1}^{m} D_{n-r}^{m+1},$$

(m, r, a_i, constants).

5737. *Proposed by Simeon Reich, Israel Institute of Technology*

Let $p(z)=a_0+a_1z+ \cdots +a_{n-1}z^{n-1}+a_nz^n$, $a_n\neq 0$, be a polynomial with

complex coefficients. Denote

$$\{\max_{0\leqq k\leqq n-1} |a_k/a_n|\}^{1/n}$$

by Q. It is known (Morris Marden, *The Geometry of the Zeros*, AMS, 1949, p. 96) that all the zeros of $p(z)$ lie in the circle

$$|z| \leqq 1 + Q^n.$$

Suppose that $a_{n-1}=0$, and that $Q>1$. Prove that the zeros of $p(z)$ lie also in the circles

$$|z| \leqq Q + Q^2 + \cdots + Q^{n-1}, \quad \text{and} \tag{1}$$

$$|z| \leqq \max_{0\leqq j,k\leqq n-1,\, j\neq k} \{(1+|a_k/a_n|)(1+|a_j/a_n|\}^{1/2}. \tag{2}$$

5738. *Proposed by E. M. Reingold, Cornell University*

Prove that there are infinitely many primes whose representations in base b begin (on the left) with an arbitrary string of digits $a_n a_{n-1} \cdots a_0$ (where $0\leqq a_i<b$).

5739. *Proposed by Richard Johnsonbaugh, Morehouse College*

Find all connected, locally compact abelian groups with connected dual group.

SOLUTIONS OF ADVANCED PROBLEMS

Theorem of the Primitive Element

5668 [1969, 423]. *Proposed by P. M. Perdew, University of Hawaii*

Prove the following generalization of the Theorem of the Primitive Element. Let $K=F(\alpha, \beta)$, where α and β are algebraic over F. Denote by r and s the multiplicities of α and β, and set $t=\min(r, s)$. Then $(t, \text{char}(F))=1$ implies K/F is a simple extension.

Solution by Robert Gilmer, Florida State University. If F has characteristic zero, the result follows from the Theorem of the Primitive Element. If $\text{char}(F)=p\neq 0$, then r and s are powers of p: hence $(t, p)=1$ implies that r or s is $p^0=1$, so that α or β is separable over F. It is then known (B. L. van der Waerden, *Modern Algebra*, vol. I, 2nd edition, Section 40) that $F(\alpha, \beta)$ is a simple extension of F.

Also solved by E. P. Del Norte, E. F. Schmeichel, and the proposer.

Polynomial Rings and Left Ore Domains

5669 [1969, 423]. *Proposed by M. E. Harris, University of Illinois in Chicago*

A *left Ore domain* R is an associative ring without zero divisors such that the intersection of any two nonzero left ideals is nonzero.

Prove: if R is a left Ore domain then so is $R[x]$ (the polynomial ring in one variable over R). How about the formal power series case?

Solution by D. Ž. Djoković, University of Waterloo, Ontario. The left Ore domain R has a left quotient ring Q which is a division ring (cf. N. Jacobson, *Theory of Rings*, p. 118). The polynomial ring $R[x]$ is canonically embedded in $Q[x]$ which is a noncommutative principal ideal domain. Hence, $Q[x]$ is a left and right Ore domain (*loc. cit.* p. 31). If $f_1(x)$ and $f_2(x)$ are elements of $R[x]$ there exist $g_1(x)$ and $g_2(x)$ in $Q[x]$ such that $g_1(x)f_1(x)=g_2(x)f_2(x)$. If $f_1(x)$ and $f_2(x)$ are nonzero then $g_1(x)$ and $g_2(x)$ can be chosen nonzero. The coefficients of $g_1(x)$ and $g_2(x)$ have the form $a^{-1}b$, where a and b are elements of R with $a\neq 0$. Since R is a left Ore domain, all the elements a have a common left multiple $m\in R$. Hence $mg_1(x)$ and $mg_2(x)$ are in $R[x]$, which proves that the intersection of the left principal nonzero ideals $R[x]f_1(x)$ and $R[x]f_2(x)$ is nonzero. The result is true for arbitrary nonzero left ideals of $R[x]$ since such ideals contain principal nonzero left ideals. This proves that $R[x]$ is a left Ore domain.

The above method does not work in the case of $R[[x]]$, the ring of formal power series in one variable over R.

Also solved by R. Berghout & R. Groenhout (Australia), Israel Kleiner, and the proposer.

Note. Berghout and Groenhout provide the reference: M. A. R. Qureshi, *(ϕ, d)—extensions of a ring*, Jour. Natur. Sci. and Math., 7 (1967) 71–83 (Math. Rev. vol. 38, #192). Kleiner provides the reference: L. Small, *Orders in Artinian rings*, Jour. Alg., 4 (1966) Theorem 2.43, where generalizations of the problem are given. However, as implied by the above, we have received no material on the major thrust of the problem as it relates to the ring of formal power series.

REVIEWS

EDITED BY KENNETH O. MAY

COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER

Printed materials for review should be sent to Kenneth O. May, American Mathematical Monthly, Department of Mathematics, University of Toronto, Toronto 5, Canada. Films and correspondence relating to films should be sent to Seymour Schuster, Carleton College, Northfield, Minnesota 55057.

All unsigned material is written by the editors. A boldface capital **C** *in the margin indicates that a review is based in part on classroom use. Professors willing to write such a review should first inform the editor in order to avoid duplication.*

A SURVEY OF MATHEMATICS TEXTS FOR ELEMENTARY SCHOOL TEACHERS

LEONARD FELDMAN, Makerere University, Uganda and San Jose State College

A well-rounded number systems course for elementary school teachers requires an exploratory spirit, a progressively increasing level of rigor, relevance to the changing school curricula, and mathematical content compatible with CUPM recommendations. Furthermore it should encourage the student to

THE AMERICAN MATHEMATICAL MONTHLY

(FOUNDED IN 1894 BY BENJAMIN F. FINKEL)

THE OFFICIAL JOURNAL OF

THE MATHEMATICAL ASSOCIATION OF AMERICA, INC.

VOLUME 77 NUMBER 6

CONTENTS

(continued on inside cover)

JUNE-JULY 1970

NOTICE TO AUTHORS

Articles should be typewritten on good quality paper, triple spaced with wide margins. Submit the original and a duplicate; keep a complete copy for protection against loss. Please follow the format used in current issues of the Monthly. The manuscript must make the author's intent clear to the printer. A paper accepted for publication by the editor but not acceptable to the printer will be returned to the author for retyping. Please use the American Mathematical Society *Manual for Authors of Mathematical Papers* as your guide and please conform to correct English usage. A *Manual for Monthly Authors* is available on request.

Prospective Monthly authors are advised to consult the Statement of Policy in the January 1969 issue, p. 2. Specialized research is usually unsuitable. Backlog: Main Articles 11 months, Math. Notes 10 months, Research Problems 7 months, Classroom Notes 7 months, Math. Education 8 months.

EDITORIAL CORRESPONDENCE AND MAIN ARTICLES: to Harley Flanders, American Mathematical Monthly, Tel Aviv University, Ramat Aviv, Israel (see Notice, p. 555); NOTES, etc.: to the corresponding Associate Editor; ADVERTISING CORRESPONDENCE: to Raoul Hailpern, Mathematical Association of America, SUNY at Buffalo, Buffalo, N. Y. 14214; CHANGE OF ADDRESS and SUBSCRIPTIONS: to A. B. Willcox, Mathematical Association of America, 1225 Connecticut Ave., N.W., Washington, D. C. 20036.

Annual dues for members of the Association (including a subscription to the American Mathematical Monthly) are $10.00. For nonmembers the subscription price is $15.00.

Published by the Association at Washington, D. C., and Menasha, Wisconsin during the months of January, February, March, April, May, June–July, August–September, October, November, December.

Second-class postage paid at Washington, D. C., and additional mailing offices. Acceptance for mailing at special rate of postage provided for in the Act of February 28, 1925, embodied in Paragraph 4, Section 538, P. L. and R., authorized April 1, 1926.

PRINTED IN THE UNITED STATES OF AMERICA

DYNAMIC TOPOLOGY

G. T. WHYBURN, University of Virginia

1. Introduction. This title is not meant to be either laudatory or derogatory but simply to be in contrast with what might be called static topology. The paper of mine which attracted the notice of the Chauvenet Prize Committee in the 1930's was called "On the structure of continua" and dealt largely with cyclic element [1] and other types of structure analysis of certain topological spaces. Even in this, however, the problem was formulated as to what type of mappings will preserve *all* topological properties, especially when applied to special domain spaces such as Euclidean spaces.

Briefly put, dynamic topology refers to the body of results of a topological nature in which the function concept plays an essential role. The function concept is basic in nearly all fields of mathematics. We study spaces and systems by observing how they act when undergoing transformations affected by the action of functions upon them. This is similar to methods used in the physical and natural sciences in studying concrete problems and phenomena of the world around us. In most cases movement, growth, change under natural or artificially induced conditions, evolution, or the like appropriate to the medium under review are involved in physical, chemical or biological studies. In mathematics the function concept enjoys an ever increasing role in nearly all areas, whether it be analysis, algebra or topology. This is especially true now and is of fairly recent vintage in topology. However, there is no lessening of interest and activity in the static study of structures. The Poincaré conjecture, for example, is the source of much fruitful effort as are also manifold structure and characterization and the relation between a set and its complement in Euclidean and other manifolds [2, 3, 4]. Even these areas of topology, however, become more and more pervaded by usage of the function device. How a space changes or remains invariant under action of various types of functions is most revealing as to its structure.

Only a few phases of this subject will be dealt with in this lecture and these are selected more because of their relevance to my own work and interests than on other grounds. However, it is hoped that they will be of interest to you also, at least in that they are significant examples of the work going on in certain areas of topological research. We will be concerned largely with results and problems related to compactness of mappings and compactifications of mappings. Thus an appropriate subtitle would be "Compactness and compactifications of mappings."

To simplify our base of discussion, unless otherwise indicated, all topological spaces will be assumed to be Hausdorff spaces, i.e., distinct points lie in disjoint open sets. A *mapping* $f: X \rightarrow Y$ of a space X into another space Y is a *continuous* function from X, its *domain*, to Y, its *range*. The double arrow $f: X \Rightarrow Y$ will indicate that the function maps X *onto* Y. The boundary $\overline{U} - U$ of an open set U is denoted by ∂U.

2. Compactness and quasi-compactness. A mapping $f\colon X \rightarrow Y$ is *compact* [5, 6] provided the inverse $f^{-1}(A)$ of every compact set A in Y is compact. Since X and Y are Hausdorff spaces, note that all mappings $f\colon X \rightarrow Y$ are compact when X is a compact space. However, the mapping $w=e^{i\theta}$ of $0 \leqq \theta < 2\pi$ onto $|w|=1$ is not compact, for example. In general a mapping is compact if it is closed and has compact point inverses. The converse holds precisely [7] when X and Y are Hausdorff spaces and Y is a k-space, i.e., when the topology in Y is determined by its compact subsets in the sense that a set in Y is closed if and only if it meets every compact set in a closed set. Assuming the above conditions, compactness of a mapping also is readily characterized in terms of filters or directed families, $f\colon X \Rightarrow Y$ being compact if and only if for every directed family N in Y converging to $y \in Y$ the inverse family $f^{-1}(N)$ is directed toward $f^{-1}(y)$ in the sense that every directed underfamily has a cluster point in $f^{-1}(y)$. Thus compactness of a mapping means precisely that compactness of sets is invariant both forward and backward, under both f and f^{-1}. It is such a useful property because it assures one of action on noncompact domain spaces closely resembling the action of a mapping on compact spaces with which so much can be done.

A mapping $f\colon X \rightarrow Y$ is *quasi-compact* [8] provided the image of every closed (open) inverse set is closed (open). This property was introduced by Alexandroff-Hopf [9] and called strong-continuity. More recently it has come to be known as the property of being a quotient mapping. It will be noted at once that every closed mapping is automatically quasi-compact, as is also every open mapping, and in particular every compact mapping, granted in the latter case that the domain and range spaces are Hausdorff and that the range is a k-space. It also readily follows that every retraction is quasi-compact. In particular note that the exponential mapping $w=e^z$ is quasi-compact but not compact. The same is true of the mapping given by any transcendental entire function.

Quasi-compactness is closely related to and is especially useful in the study of mappings generated by decompositions of a given space X. Indeed if X is decomposed into a collection G of disjoint sets $\{g\}$ and a new space Y' is set up with elements of G as its points and a topology introduced by defining a set in Y' to be open if and only if the union of its elements is an open set in X, the so-called quotient topology, then the natural mapping $\phi\colon X \Rightarrow Y'$ which maps each $x \in X$ into the element g of G containing x is quasi-compact. If we begin with a mapping $f\colon X \Rightarrow Y$ and let X be decomposed into the collection $\{f^{-1}(y)\}$ of point inverses, $y \in Y$, then let $\phi\colon X \Rightarrow Y'$ be the natural mapping of this decomposition, the function $h\colon Y' \Rightarrow Y$ defined by $h(y')=f\phi^{-1}(y')$ is 1-1 and continuous. Further h^{-1} will be continuous and thus h will be a homeomorphism if and only if f is quasi-compact.

Openness and closedness of a mapping are nicely characterized in terms of quasi-compactness as follows: A mapping $f\colon X \Rightarrow Y$ is open (closed) if and only if it is quasi-compact and generates a lower (upper) semi-continuous decomposition of X into point inverses. Another remarkable property of quasi-compact mappings is that local connectedness is invariant under all such mappings,

thus under all open or closed mappings and all retractions. The latter results from the unexpected fact that a mapping $f: X \Rightarrow Y$ quasi-compact on any cross section (i.e., any set in X mapping onto Y) is automatically quasi-compact on X.

The inverse does not hold, however. Indeed, in contrast to the situation with open or closed mappings, it does not follow [10] that a quasi-compact mapping is quasi-compact when restricted to an inverse set—not even to its kernel [$=$the set of all $x \in X$ for which $x = f^{-1}f(x)$]. This fact has important implications in the case of invariance of connectedness under the inverse f^{-1} of a mapping $f: X \rightarrow Y$. In case point inverses are connected then X will be connected if Y is connected, provided f is quasi-compact. However, it does not follow from this that the restriction $f|I$ of f to an inverse set I will enjoy the same property, because this restriction may fail to be quasi-compact. What is needed here is that f be hereditarily quasi-compact, i.e., quasi-compact on every inverse set. It turns out that this latter property can be assured for *all* quasi-compact mappings onto a given range space Y, provided Y is an *accessibility space* [11]. A regular Hausdorff space Y is an accessibility space provided each limit point p of a set M in Y is *accessible by closed sets* in Y in the sense that $M+p$ contains a closed set C having p as a limit point. Any quasi-compact mapping onto a regular accessibility space is automatically hereditarily quasi-compact. Thus, in particular, if point inverses for such a mapping are connected then connectedness is fully invariant under the inverse. The notion of accessibility space is readily extended to spaces other than regular. Thus it may be said that for mappings with connected point inverses, connectedness is invariant both forward and backward, provided the mapping is hereditarily quasi-compact and this is true for all quasi-compact mappings into accessibility range spaces. In particular it holds for such mappings into locally compact or first countable range spaces as these are always accessibility spaces. We may summarize by saying: under a compact mapping compactness is invariant both forward and backward; and under a quasi-compact weakly monotone mapping onto an accessibility space, connectedness is invariant both forward and backward.

3. Monotoneity. We next consider compactness and quasi-compactness of a mapping and the compact part of a mapping in a somewhat more restricted setting involving local compactness of the domain and monotoneity of the mapping. A mapping $f: X \rightarrow Y$ is *monotone*, provided its point inverses $f^{-1}(y)$ are continua, i.e., compact connected sets. The sequence of results which we indicate next leads to interesting applications for mappings on the Euclidean spaces.

(1) *If X is locally compact, every quasi-compact monotone mapping $f: X \Rightarrow Y$ is closed (and thus compact) and its range Y is also locally compact.*

To prove this, take any closed set K in X and any $p \in Y - f(K)$, and let $P = f^{-1}(p)$. Then take an open set U about P with $\overline{U}$ compact and $U \cdot K = \emptyset$, using compactness of P and the fact that X is a Hausdorff space. Then ∂U and $C = f(\partial U)$ are compact and thus closed. Let $W = Y - C$ and $R = U \cdot f^{-1}(W)$. Then

R is open, $P \subset R \subset U$, and $R = f^{-1}f(R)$ because $f^{-1}(y)$ is connected and meets R for each $y \in f(R)$ but cannot meet ∂U. Accordingly $f(R)$ is open and contains p but does not meet $f(K)$. Thus $f(K)$ is closed. That Y is locally compact follows from the invariance of local compactness under closed mappings.

(2) *If X and Y are noncompact but locally compact spaces, a mapping $f: X \Rightarrow Y$ is compact if and only if it has a continuous extension f^* from the 1-point compactification X^* of X to the 1-point compactification Y^* of Y.*

The extension is unique when it exists. To prove (2) we suppose f is compact. Let $X^* - X = x^*$, $Y^* - Y = y^*$ and define $f(x^*) = y^*$. Let V be any open set in Y^*. If V contains y^*, $Y^* - V = Y - V$ and this set is compact. Accordingly

$$f^{-1}(Y - V) = X^* - f^{-1}(V) = X - f^{-1}(V)$$

and this set is compact since $f|X$ is compact. Hence $X^* - f^{-1}(V)$ is closed so that $f^{-1}(V)$ is open in X^*. If V doesn't contain y^*, $f^{-1}(V)$ is open in X and thus also in X^*. Hence the extension of f to X^* is continuous.

On the other hand, suppose f has a continuous extension $g: X^* \rightarrow Y^*$, so that $f = g|X$. Then since X^* and Y^* are compact but Y is noncompact, g must be an onto mapping and it obviously is compact. Thus f is compact, because X is an inverse set for g and any restriction of a compact mapping to an inverse set is compact.

(2′) COROLLARY. *If X and Y are locally compact, a monotone mapping $f: X \Rightarrow Y$ is compact if and only if it has a continuous monotone extension from X^* to Y^*.*

Proof for Corollary (2′). In case neither X nor Y is compact the proof for (2) shows that when f is compact its extension to X^* is monotone when f is monotone, and that when f has a continuous extension to X^*, f is automatically compact.

There remains the case in which X or Y is compact. If X is compact, Y is necessarily compact and all mappings from X to Y are compact and all have trivial extension to X^*, which is the same as X. If Y is compact, then for $f: X \Rightarrow Y$ compact, $X = f^{-1}(Y)$ is compact, and again f is trivially extendable to X^*. On the other hand, still assuming Y compact, if f has a monotone extension g to X^* and if X^* were different from X, say $X^* - X = x^*$, we would have $g(x^*) = y_0 \in Y$ and $g^{-1}(y_0) = x^* + f^{-1}(y_0)$, a disconnected set, contrary to monotoneity of g.

Monotoneity is essential in (2′). For if f is the 1-1 mapping of a plane (or punctured sphere) obtained by identifying the two poles of a sphere and then deleting one of the poles from the domain of f, then f has a continuous but nonmonotone extension to its 1-point compactification (the sphere) but f is not compact of course.

(2″) COROLLARY. *Every rational function $w = f(z)$ generates a compact mapping of the z-plane onto the w-plane.*

For any function $f: X \Rightarrow Y$ and any subset Y' of Y, any subset X' of X

mapping *onto* Y' is called a *trace* of Y' under f. Note that a trace of Y itself is the same as a *cross section* of f as used earlier in this discussion.

(3) THEOREM. *If $f: X \Rightarrow Y$ is a monotone mapping where X is locally compact, then any set H in Y which has a compact trace K has a compact inverse* [**12**].

We set $N = f^{-1}(H)$ and consider the restriction $f|N = g$ of f to N. Since K is compact the mapping $g|K: K \Rightarrow H$ is compact and thus, in particular, quasi-compact. Since K is a cross section for the mapping $g: N \Rightarrow H$, it follows that g is quasi-compact. Since g is also monotone it follows by (1) that g is compact. Accordingly the set $g^{-1}(H) = N = f^{-1}(H)$ is compact.

(3′) COROLLARY. *If X is locally compact and $f: X \Rightarrow Y$ is a monotone mapping, for any compact set K in X, $f^{-1}f(K)$ is compact. If K is a continuum so also is $f^{-1}f(K)$.*

(3″) COROLLARY. *A monotone mapping on a locally compact space is compact if and only if each compact set in Y has a compact trace.*

It may be noted that the monotoneity condition in this theorem is quite essential, even though point inverses are compact. For if X is the part of the graph of the parabola $y^2 = x$ for which $-1 < y \leqq 1$, the vertical projection of X onto the interval $I = [0, 1]$ meets all conditions except monotoneity and yet I has a compact trace but a noncompact inverse.

4. An application. Invariance of the plane. One of the most outstanding results in the field of topology asserts that if we decompose the Euclidean plane into an upper semi-continuous collection of disjoint continua, not separating the plane, the decomposition space is itself a topological plane and thus homeomorphic with the original plane. This theorem (R. L. Moore [**13**]) still has no satisfactory simple extension to higher dimensional Euclidean spaces. We indicate next how a form of this fine theorem may be obtained from the sequence of results just discussed.

For the purposes of this discussion we define an $\mathcal{S}^2$ (2-sphere) as a nondegenerate unicoherent Peano continuum containing a simple closed curve and separated irreducibly by every such curve it contains. Similarly an $\mathcal{E}^2$ is any locally compact Hausdorff space whose 1-point compactification is an $\mathcal{S}^2$ or, equivalently, which is homeomorphic with the complement of some point on some $\mathcal{S}^2$. That any $\mathcal{S}^2$ or $\mathcal{E}^2$ is homeomorphic with an ordinary 2-sphere or plane respectively will not be used or proved in this discussion, although this is the case as is well known. Indeed the condition of unicoherence is redundant and imposed only for convenience in this discussion.

We first sketch briefly a proof for

(4) THEOREM. *The property of being an $\mathcal{S}^2$ is invariant under nonconstant, nonseparating monotone mappings onto Hausdorff spaces.*

It is necessary first to establish the following sequence (i)–(v) of results in any $\mathcal{S}^2$.

(i) JORDAN CURVE THEOREM: Every simple closed curve J on $\mathcal{S}^2$ separates $\mathcal{S}^2$ into exactly two components and is the boundary of each of these. The only thing left to prove here is that there are just two components of $\mathcal{S}^2-J$. This is accomplished by constructing an auxiliary simple closed curve J' using two disjoint arcs of J and arcs in two of the components of $\mathcal{S}^2-J$. Existence of a third component of $\mathcal{S}^2-J$ would force all of J into the closure of a single component of $\mathcal{S}^2-J'$ so that J' could not separate $\mathcal{S}^2$.

(ii) θ-CURVE THEOREM: For any θ-curve θ on $\mathcal{S}^2$, $\mathcal{S}^2-\theta$ has exactly three components and each of these is bounded by a unique pair of the edges of θ. This is an easy consequence of the Jordan Curve Theorem.

(iii) BOUNDARY CURVE THEOREM (Torhorst): The boundary B of any component R of the complement of a Peano continuum M on $\mathcal{S}^2$ is itself a Peano continuum. This is proven by showing that B is a continuum by unicoherence of $\mathcal{S}^2$ and, if it is not locally connected, a θ-curve θ can be constructed in M so that each of its 3 open edges can be connected to B without meeting either of the other two edges—thus contradicting the fact that R lies in just one of the three complementary regions of θ on $\mathcal{S}^2$.

(iv) SEPARATION THEOREM: If A and B are continua on $\mathcal{S}^2$ so that $A\cdot B$ is totally disconnected and both $A_0=A-A\cdot B$ and $B_0=B-A\cdot B$ are connected, then $\mathcal{S}^2$ contains a simple closed curve J with $J(A+B)=A\cdot B$ which separates A_0 and B_0 in $\mathcal{S}^2$. Further, if B does not separate $\mathcal{S}^2$, then, for any $\sigma>0$, J can be chosen so that the component of $\mathcal{S}^2-J$ containing B_0 lies in the σ-neighborhood of B.

This is proven by covering B_0 with the union of a sequence of regions whose closures are Peano continua not meeting A and so that only finitely many are not in the ϵ-neighborhood of $A\cdot B$ for each $\epsilon>0$. Then by unicoherence of $\mathcal{S}^2$, $\overline{Q}$ is a Peano continuum. It then follows that the boundary J of the component R of $\mathcal{S}^2-\overline{Q}$ containing A_0 meets all our requirements. (Note that in case B doesn't separate $\mathcal{S}^2$, for a given σ we may suppose A_0 contains all points of $\mathcal{S}^2$ not in the σ-neighborhood of B.)

To see this we first note that J must be either an arc or a simple closed curve. For if J contains a simple closed curve C, a simple θ-curve construction along with (ii) shows that J must reduce to C. On the other hand if J were a dendrite it could have no branch point because again a θ-curve construction in $J+R$ would likewise lead to a contradiction of (ii).

The proof is then completed by applying the part of the conclusion already established to show that J cannot be a simple arc because its end points would be accessible from R so that J would lie on a simple closed curve in $\mathcal{S}^2$ and hence could not separate $\mathcal{S}^2$. That any point p of J would be accessible from R follows from the fact that $R+p$ is locally connected because, by the part already proven, p lies in an arbitrarily small region in $\mathcal{S}^2$ bounded by an arc or a simple closed curve C which meets J in at most two points so that $R\cdot C$ would have at most three components.

(v) DYGON THEOREM. A dygon is a continuum D which is the union of two continua A and B meeting in exactly two points x and y (the vertices of D) and so that $A_0 = A-(x+y)$ and $B_0 = B-(x+y)$ (the edges of D) are connected sets.

The Dygon Theorem asserts that: *Every dygon D on $\mathcal{S}^2$ separates $\mathcal{S}^2$; and if the boundary of any component of $\mathcal{S}^2-D$ meets both edges of D it also contains both vertices of D.*

The proof is an easy application of (iv) and (ii). For (iv) yields a simple closed curve J in $\mathcal{S}^2$ separating the edges of D and meeting D in just its vertices x and y. The two open arcs of J from x to y are then separated in $\mathcal{S}^2$ by D, since otherwise one can easily construct a θ-curve containing J and meeting D only in $x+y$. This contradicts (ii).

The Dygon Theorem yields the following key lemma:

LEMMA. *Let $\phi: S \Rightarrow \Sigma$ be monotone, where S is an $\mathcal{S}^2$. Any simple closed curve J in Σ containing two distinct points x and y with unique inverses separates Σ; and if the boundary of any component of $\Sigma - J$ meets both open arcs xy of J, it must also contain both x and y.*

Proof of Lemma. Since ϕ is monotone, $\phi^{-1}(J)$ is a dygon D with vertices $x' = \phi^{-1}(x)$ and $y' = \phi^{-1}(y)$. Hence D separates S so that J separates Σ by monotoneity of ϕ. Likewise, if Q is any component of $\Sigma - J$ whose boundary B meets both open arcs xy of J, the boundary of $\phi^{-1}(Q)$ meets both edges of D and thus contains both x' and y', so that B necessarily contains both x and y.

With this lemma we are now in position to prove (4), the Sphere Invariance Theorem:

The property of being an $\mathcal{S}^2$ is invariant under nonconstant monotone nonseparating mappings onto a Hausdorff space.

Proof. Let $f: S \rightarrow \Sigma$ be such a mapping where S is an $\mathcal{S}^2$. Then Σ is a locally connected unicoherent continuum since these properties are invariant under monotone mappings. Also Σ has no cut point since f is nonseparating, and thus Σ contains a simple closed curve.

Thus if Σ is not an $\mathcal{S}^2$ it must contain a simple closed curve J such that $\Sigma - J$ either (1) is connected or (2) contains a component whose boundary is an arc ab of J. Let x and y be distinct points of J, chosen so that they separate a and b on J in case (2) holds.

Decompose S into the continua $f^{-1}(x)$, $f^{-1}(y)$ and individual points of $S - f^{-1}(x) - f^{-1}(y)$. Let $\phi: S \Rightarrow S'$ be the natural mapping of this decomposition. Then ϕ is monotone, nonconstant and nonseparating, so that S' is a unicoherent locally connected continuum containing a nondegenerate simple closed curve. Also, if J' is any simple closed curve in S', and a_1 and b_1 are any two points of J', there always exist points x_1 and y_1 with unique ϕ inverses so that x_1 and y_1 separate a_1 and b_1 on J'. Hence, by the lemma, J' separates S', and no simple arc ab of J' can separate S'. Accordingly S' is an $\mathcal{S}^2$.

Now define $g: S' \Rightarrow \Sigma$ by $g(p) = f\phi^{-1}(p)$ for $p \in S'$. It is verified at once that g

is monotone and nonseparating. Thus since $g^{-1}(x)$ and $g^{-1}(y)$ are single points of S', the lemma applied to g shows (a) that J separates Σ so that case (1) cannot arise and (b) that there can be no component of $\Sigma - J$ whose boundary is an arc ab of J as required by case (2). Thus the supposition that Σ is not an $\mathcal{S}^2$ leads to a contradiction.

As an easy consequence we now get

(5) THEOREM. *The property of being an $\mathcal{E}^2$ is invariant under quasi-compact monotone nonseparating mappings onto Hausdorff spaces.*

For let $\phi: X \Rightarrow Y$ be such a mapping where X is an $\mathcal{E}^2$. Then ϕ is closed and Y is locally compact by (1). Thus by (2), ϕ has a continuous extension $\phi^*: X^* \Rightarrow Y^*$ to the 1-point compactification X^* of X onto that of Y. Then X^* is an $\mathcal{S}^2$ and ϕ^* is nonconstant, nonseparating and monotone and Y^* is a Hausdorff space. Thus by (4), Y^* is an $\mathcal{S}^2$; and hence Y itself is an $\mathcal{E}^2$.

REMARK. In view of the well-known fact, not proven here (see above), that all $\mathcal{S}^2$'s are homeomorphic with each other, as are also all $\mathcal{E}^2$'s, as a consequence of (4) and (5) we have the classical result of R. L. Moore which asserts that:

The decomposition space for any nontrivial upper semicontinuous decomposition of the plane, or 2-sphere X into disjoint continua not separating X, is homeomorphic with X.

5. The compact part of a mapping. For any mapping $f: X \Rightarrow Y$ we define (i) Q as the union of the interiors of sets in Y having a compact inverse and $P = f^{-1}(Q)$, (ii) Q' as the union of the interiors of the images of all compact sets in X. It results at once that P, Q and Q' are open sets, and that the mapping $f|P: P \Rightarrow Q$ is compact. Also, using Theorem (3) above, we get at once

THEOREM A. *Let X be locally compact. If f is monotone then $Q = Q'$. If f is* (1-1), *Q is exactly the set of points of continuity of f^{-1}.*

THEOREM B. *Let X be locally compact and have a countable base and suppose Y is a complete metric space. Then Q' is dense in Y. Further, if f is monotone $Q = Q'$; and if f is* 1-1, *f maps P topologically onto Q.*

6. One-to-one Mappings. These are always of special interest. One is naturally concerned with conditions under which they are homeomorphisms and with the extent to which they preserve the topological structure of their domain spaces even when they are not homeomorphisms.

We have already noted that $w = e^{i\theta}$, $0 \leqq \theta < 2\pi$, maps the half open interval onto the circle $|z| = 1$ in 1-1 fashion and also that the plane can be mapped similarly onto the pinched 2-sphere obtained by identifying its poles. In both of these cases the range space is compact and the domain locally compact. Of course if the domain space is compact, the mapping is a homeomorphism. Also any quasi-compact 1-1 mapping is a homeomorphism.

In general the 1-1 continuous image of even a simple noncompact space may be quite different in structure from that of the original. For example such

an image of the line may fail to be locally connected, as is well known. If we consider only 1-1 mappings onto locally connected spaces, significant results have been found by McAuley and Lelek [14] and Jones [15] in case the domain is a line or a plane. The only possible 1-1 locally connected continuous images of the line, for example, are the line, figure eight, noose, dumbbell or θ-curve.

In this particular case it may be noted that if we impose the additional condition of unicoherence on the image space, then the only possible image of the line is the topological line itself. Further, in this case, the mapping itself must be topological. This and other related observations and results led to the conjecture at one time that a 1-1 mapping from a locally connected locally compact connected separable metric space onto a unicoherent space of the same sort would necessarily be a homeomorphism. Indeed such a result was asserted [16] and thought to be proven at one time. However, examples were given by Kenneth Whyburn [17] and L. C. Glaser [18] showing that this is not true and, indeed, not even when the image space is E^n for $n \geqq 3$. The case of 1-1 mappings of such spaces onto E^2 has been studied by E. Duda [19] who has shown that these mappings are homeomorphisms provided the domain spaces satisfy some simple additional restrictions.

A fairly simple example in which the image is a unicoherent Peano continuum (but not a Euclidean space) may be constructed as follows. Let $\triangle$ be an equilateral triangle ABC including its interior and let O be the center of $\triangle$. Then add to $\triangle$ three solid triangles $A'OB$, $B'OC$, and $C'OA$ meeting $\triangle$ in just the segments OB, OC, OA respectively, and meeting each other by pairs only in O. Let X' be the union of these 4 solid triangles, and let X' be mapped onto a space Y by identifying A' with A and mapping $A'B$ linearly onto AB; similarly $B'C$ is mapped onto BC and $C'A$ onto CA. Otherwise the mapping is 1-1. Thus Y consists of $\triangle$ plus three open pockets created by identifying points on $A'B$ with corresponding points on AB, and similarly for $B'C$ and $C'A$. Finally, we delete from the domain X' of the mapping all points on the periphery of the original triangle $\triangle = ABC$. Let X be the resulting space $X = X' -$ (periphery of $\triangle$), and $f: X \Rightarrow Y$ the restriction of the mapping to X. Then f is 1-1 and continuous, X is locally compact, connected and locally connected, and Y is a unicoherent Peano continuum, but of course f is not a homeomorphism. (Indeed, Y is compact but X is not.)

It is not enough, in general, to assume X and Y homeomorphic with each other in order to assure that every 1-1 mapping of X onto Y will be a homeomorphism, even when these spaces are locally connected generalized continua ($=$locally compact, connected separable metric spaces). It is true, however, that in case both X and Y are Euclidean spaces, any 1-1 mapping of X onto Y must be a homeomorphism. This conclusion is obtainable as a consequence of the Brouwer Theorem on invariance of openness in Euclidean spaces. Indeed we can get a somewhat better result by formulating the Brouwer Property in a broader setting as follows.

A space X is said to have the Brouwer Property [12], provided any subset of X which is homeomorphic with some open set in X is itself open in X. All Euclid-

ean spaces have this property by Brouwer's Theorem, as do also all Euclidean manifolds. However, a large class of spaces which are not Euclidean also have this property. For example, the set obtained by taking the 1-point compactification of a surface of infinite genus. Also the generalized closed manifolds of Wilder have the Brouwer Property although they are not all Euclidean by any means. A straightforward argument suffices to prove

THEOREM. *If X and Y are homeomorphic locally compact spaces having the Brouwer Property, any* 1-1 *mapping of X onto Y is a homeomorphism.*

Now in case X and Y are Euclidean spaces one of which is mappable onto the other by a 1-1 mapping, it results from Theorem B in section 5 above that X and Y are of the same dimension and thus are homeomorphic with each other. Thus any 1-1 mapping of one Euclidean space onto another is a homeomorphism.

For further examples and results in connection with the Brouwer Property the reader is referred to a study made by E. Duda [**20**].

7. The compactness problem. Results of this sort in the 1-1 case led the author many years ago to formulate and study the problem of determining conditions under which a mapping of a Euclidean space E^n onto itself is necessarily compact. The cases $n=1, 2$, were solved and appeared in a paper published in 1959 [**12**]. For $n=1$ we need only assume compactness of point inverses and for $n=2$ it is enough to assume them to be continua, i.e., assume the mapping monotone. Higher dimensional cases of the problem present considerably greater difficulty. We now consider these various cases in some detail.

(i) $n=1$. *If X and Y are lines (E^1), a mapping $f:X\Rightarrow Y$ is compact if (and only if) it has compact point inverses.*

Let f have compact point inverses and take any compact set K in Y. Let ab be an interval in X whose interior contains $f^{-1}(\alpha+\beta)$, where $\alpha\beta$ is an interval in Y whose interior contains K. Then neither $-\infty a$ nor $b\infty$ can meet $f^{-1}(K)$. For if, say, $-\infty a$ meets $f^{-1}(K)$, $f(-\infty a)$ lies wholly in $\alpha\beta$ so that $f(-\infty a+ab)$ is bounded. Thus $f(b\infty)$ would meet both $-\infty\alpha$ and $\beta\infty$ and thus would contain α and β which is impossible.

In terms of real valued functions this result says that if f is a real function of a real variable continuous everywhere on E^1 which takes each real value at least once, but only on a compact set, then

$$\lim_{x\to\infty} f(x) = \pm\infty, \quad \lim_{x\to-\infty} f(x) = \mp\infty.$$

(ii) $n=2$. In the case of a mapping of a plane onto a plane it is not enough to have compact point inverses in order to make the mapping compact. For the function

$$w = \left(\frac{x}{x+1} + iy\right)^2,$$

i.e., $w = z'^2$, where

$$z' = \frac{x}{x+1} + iy,$$

maps the complex plane onto itself and yet f takes on each value either one or two times. However, as indicated earlier, it is sufficient (though not necessary, of course) that point inverses be continua.

THEOREM. *If X and Y are planes, any monotone mapping $f: X \Rightarrow Y$ is compact (and nonseparating).*

Proof. The map f generates a decomposition of X into continua. Let Y' denote the decomposition space and $\phi: X \Rightarrow Y'$ the decomposition map. There is the one-to-one map $h: Y' \Rightarrow Y$ with $h\phi = f$. Now ϕ is quasi-compact, hence compact by (1) of section 3. We therefore have a unique monotone map $\phi^*: X^* \Rightarrow Y'^*$ as in (2) of section 3. By the general form of Moore's Theorem, Y'^* is a cactoid.

Case 1: Y'^* is a sphere. In this case Y' is a plane. Hence, the one-to-one map $h: Y' \Rightarrow Y$ must be open and a homeomorphism by Brouwer's Theorem. Hence f is compact in this case.

Case 2: Y'^* contains a sphere as a proper subset. No sphere can map one-to-one into the plane, hence Y' must contain a plane E which has limit points of $Y' - E$. Then h maps E homeomorphically onto an open subset of Y. Since E contains limit points of $Y - E$, it is impossible to extend $h: E \rightarrow Y$ to a continuous one-to-one map of Y' into Y.

Case 3. Y'^* is a dendrite. By Theorem B, h must be a homeomorphism of some nonempty open subset onto an open subset of Y. But every open subset of a dendrite has cut points, whereas no open subset of the plane has cut points.

Hence f is compact. The mapping $f^*: X^* \Rightarrow Y^*$ is clearly nonseparating, hence so also is $f: X \Rightarrow Y$.

Since the mapping f in this theorem turns out to be automatically nonseparating, in view of results on invariance of the plane discussed earlier we have the

COROLLARY. *Let $f: X \Rightarrow Y$ be a monotone mapping where X is a plane and Y is a Hausdorff space. Then Y is a topological plane if and only if f is compact and nonseparating.*

(iii) $n \geqq 3$. These cases were discussed by the author at the Georgia Conference on Topology of 3-manifolds in 1961 (see Proceedings of this Conference, by M. K. Fort, Prentice-Hall, 1961). It was conjectured that a 1-monotone mapping of E^3 onto E^3 is always compact and, possibly, an $(n-2)$-monotone mapping of E^n onto E^n is compact—a mapping being *r-monotone* provided its point inverses have trivial homology groups in dimension $\leqq r$. The conjecture in this form would reduce to exactly the above result in (ii) for $n=2$ and to the conjectured one for $n=3$. This problem and conjecture attracted considerable attention and effort on the part of those in attendance at the conference and others who learned of it later. It was shown by Connell at that time that acyclic maps

of E^n onto E^n are always compact, where "acyclic" means that homology groups of *all* dimensions of point inverses are trivial. See the above mentioned Proceedings.

In 1965 a positive solution to the problem, verifying the conjecture for general n in the form given above, was published by Väisäla, along with a number of related results. Methods of proof required, however, are much more algebraic and less elementary than needed in the cases $n=1$, 2. An independent proof of this same general result: *$(n-2)$-monotone mappings of E^n onto E^n are compact* was found by Conner and Jones and distributed in manuscript form. Their methods differ from those of Väisäla in that they depend more directly on homology theory and exact sequences rather than on the topological index. However, both proofs are far from elementary; and a simple direct set-theoretic type of proof for this fine theorem for general n would be highly desirable and would represent a valuable contribution to mathematical knowledge.

The question as to the existence of noncompact monotone ($=0$-monotone) mappings of E^n onto E^n for $n \geqq 3$ has been studied in detail by Glaser [**18**]. He has shown that such maps do indeed exist for $n \geqq 4$ and has further results on existence of monotone maps of E^k onto E^n, where k and n are not necessarily equal. At the 1969 Georgia Conference, R. H. Bing gave an example of a monotone mapping of E^3 onto E^3 which is not compact.

Also it would be of interest to determine just what properties of Euclidean spaces are essential in making these results on compactness valid. Can one formulate sufficient conditions on a space, perhaps in terms of its homology or homotopy groups or contractibility or retractability into subsets which will suffice to yield similar results on compactness to the ones discussed here for Euclidean spaces? Possibly the Brouwer Property could be useful in this connection.

8. Compactification of mappings. The existence of various forms of compactifications of a topological space X (i.e., the topological imbedding of X in a larger compact space in which X is dense) such as the one-point compactification, the Stone-Čech compactification, etc., leads naturally to the question as to whether certain types of mappings can be compactified in some similar sense. This question arose naturally also in connection with a result of Vainstein [**5**] to the effect that *any closed mapping has a partial mapping* (restriction) *which is compact and in which the image space is the same as the image of the original domain space.* The usual dual relationship between open and closed (sets, mappings, etc.) led the author to anticipate that open mappings may be related in some analogous way to compact mappings. More particularly, it was anticipated that if the domain space were suitably augmented, the mapping could be extended so as to become compact, so that the given mapping would be exhibited as a partial mapping of a compact one.

This was found to be fully correct. In a paper published in 1953 [**21**] it was shown that in a suitably constructed space unifying both the domain and range spaces, a compact mapping can be defined which is topologically equivalent to

the given one on the prototype of the original domain space. This turned out to be true for arbitrary mappings, not just for open ones. Thus it was shown that any mapping from one Hausdorff space to another is topologically equivalent to a partial mapping of a compact one (actually a retraction) in the unified space. The extended mapping is open when the original one was open; and the unified space is separable and metrizable when the given spaces are locally compact, separable and metric. More recently H. Bauer [22] had occasion to use such a result in his studies on measure preservation of mappings and, independently, he developed and applied almost exactly the same extension to show that any mapping between locally compact spaces is a partial mapping of a conservative one.

Sketched briefly, we begin with an arbitrary mapping $f:X\to Y$, where X and Y are disjoint Hausdorff spaces, and define the *unified space* Z to consist of all points in the union of X and Y with a topology in which a set $Q\subset Z$ is open if and only if

(i) *$Q\cdot X$ and $Q\cdot Y$ are open in X and Y respectively, and*
(ii) *for any compact set $K\subset Q\cdot Y$, $f^{-1}(K)\cdot(X-Q)$ is compact.*

It turns out that the injection mappings of X and Y into Z are open and closed respectively and thus are homeomorphisms. Further, if we define $r(x)=f(x)\in Z$ for $x\in X$ and $r(x)=x\in Z$ for $x\in Y$, then r is continuous and compact and hence is a compact retraction of Z onto Y. Since $r|X$ is f followed by the injection of Y into Z it is therefore topologically equivalent to f. If f is an open mapping, so also is r. If X and Y are locally compact, Z is locally compact and Hausdorff; and, in this case, if X and Y are also separable and metric so also is Z.

In general, the set X is not necessarily dense in the space Z, i.e., $\bar{X}\neq Z$ unless f is everywhere noncompact, where $\bar{X}$ denotes the closure of X in Z. However, X is dense in $\bar{X}$ of course and the partial mapping

$$r\,|\,\bar{X}:\bar{X}\to Y$$

is a compact extension of f to $\bar{X}$. Note also that since r is a retraction and $\bar{X}-X\subset Y$, $r|(\bar{X}-X)$ is the identity mapping. The original mapping, considered as a mapping of X into its image, i.e., $f:X\to f(X)$, is a compact mapping if and only if both r and r^{-1} reduce to the identity on the set $\bar{X}-X$. The important special case of a 1-1 mapping $f:X\to Y$ which is compact relative to $f(X)$, i.e., in case f is a homeomorphism of X onto $f(X)$, is of special interest here. For if f is a homeomorphism of X onto $f(X)$, it turns out that r extends f to a homeomorphsim of $\bar{X}$ onto $\overline{f(X)}$.

The unified space for a mapping has been further studied by R. Dickman [23] who developed, in particular, conditions for unicoherence of this space and implications of its unicoherence so far as the mapping is concerned.

The process of compactifying a mapping has been studied from a more general viewpoint by Dickman [24], Cain [25], and others, and quite interesting

results obtained. A compactification of a mapping $f: X \Rightarrow Y$ is defined as a pair (X^*, f^*) where X^* is a Hausdorff space containing X as a dense subspace and $f^*: X^* \Rightarrow Y$ is a closed continuous extension of f from X^* onto Y having compact point inverses. Thus f^* is a compact mapping and $f^* | X = f$.

Dickman [26] has developed the concept of maximal and minimal compactifications of spaces and, correspondingly, of mappings. Cain [27] has developed two different procedures for characterizing all possible compactifications for a given mapping. In each case there is associated with each possible compactification X' of the domain space X a compactification of the mapping $f: X \Rightarrow Y$ in a unique way so that the mapping compactification is determined by the space compactification and so that every possible mapping compactification is obtainable by the given process from a space compactification.

The first of these procedures employs filter space compactifications of spaces as developed by Wagner. The second uses in a similar way the ring compactifications of spaces provided by the structure space of the subring R of the ring of all bounded continuous real valued functions on the domain space X. This structure space has as points the maximal ideals in R and is a Hausdorff compactification of X. (See Wagner [28].)

References

1. B. L. McAllister, Cyclic elements in topology, A history, this MONTHLY, 73 (1966) 337–350.

2. R. H. Bing (editor), Summer Institute on Set Theoretic Topology, Madison, 1955.

3. M. K. Fort, Jr. (editor), Topology of 3-manifolds and related topics, Proc. University of Georgia Conference, Prentice-Hall, 1962.

4. R. L. Wilder, Topology of manifolds, A.M.S. Colloq. Publ., 32 (1949).

5 I. A. Vainstein, On closed mappings, Moskov. Gos. Univ. Uc. Zap. 155, Mat., 5 (1952) 3–53.

6. G. T. Whyburn, Open mappings on locally compact spaces, Mem. Amer. Math. Soc., 1 (1950) 24.

7. ———, Directed families of sets and closedness, Proc. N.A.S. no. 3, 54 (1965) 688–692.

8. ———, Open and closed mappings, Duke Math. J., 17 (1950) 69–74.

9. P. Alexandroff and H. Hopf, Topologie, vol. 1, Springer-Verlag, 1935.

10. G. T. Whyburn, Mappings on inverse sets, Duke Math. J., 23 (1956) 237–240.

11. ———, Notes on general topology and mapping theory, mimeographed notes, U. of Virginia, 1965.

12. ———, Compactness of certain mappings, Amer. J. Math., 81 (1959) 306–314.

13. R. L. Moore, Concerning upper semi-continuous collections of continua, Trans. Amer. Math. Soc., 27 (1925) 416–428.

14. A. Lelek and L. F. McAuley, On hereditarily locally connected spaces and one-to-one continuous images of a line, Colloq. Math., 17 (1967) 319–324.

15. F. Burton Jones, On a plane one-to-one map of a line, Colloq. Math., 19 (1968) 231–233.

16. V. V. Proizvolov, One-to-one mappings onto metric spaces, Dokl. Akad. Nauk SSSR, 158 (1964) 788–789 and 1286–1287.

17. Kenneth Whyburn, A non-topological 1-1 mapping onto E^3, Bull. Amer. Math. Soc., 71 (1965) 533–537.

18. L. C. Glaser, Dimension lowering monotone non-compact mappings of E^n, Fund. Math., 58 (1966) 177–181.

19. Edwin Duda, A theorem on one-to-one mappings, Pac. J. Math., 19 (1966) 253–257.

20. ———, Brouwer property spaces, Duke Math. J., 30 (1963) 647–660.

21. G. T. Whyburn, A unified space for mappings, Trans. Amer. Math. Soc., 74 (1953) 344–350.

22. H. Bauer, Konservative Abbildungen lokalkompacter Räume, Math. Ann., 138 (1959) 398–427.

23. R. F. Dickman, Unified spaces and singular sets of mappings of locally compact spaces, Fund Math., 62 (1968) 103–123.

24. ———, On closed extensions of functions, Proc. N.A.S., 62 (1969) 326–332.

25. G. L. Cain, Jr., Compact and related mappings, Duke Math. J., 33 (1966) 639–645.

26. R. F. Dickman, Compactifications and real compactifications of arbitrary topological spaces, to appear.

27. G. L. Cain, Jr., Compactifications of mappings, Notices, Amer. Math. Soc., 16 (1969) 277.

28. F. J. Wagner, Notes on compactification I, Nederl. Akad. Wetensch. Proc. Ser. A, 60 (1957) 171–176.

HAMILTONIAN MECHANICS AND GEOMETRY

SAUNDERS MAC LANE, University of Chicago

1. Introduction. Let me begin by thanking the organizers of the Chauvenet Symposium for the opportunity to talk at the dedication of Chauvenet Hall at the U. S. Naval Academy. It is a pleasure to honor the long tradition of devotion to Mathematics at the Academy, and to recognize again the impetus which William Chauvenet gave, in his teaching and writing, to the importance of effective exposition. Mathematical ideas do not live fully till they are presented clearly, and we never quite achieve that ultimate clarity. Just as each generation of historians must analyse the past again, so in the exact sciences we must in each period take up the renewed struggle to present as clearly as we can the underlying ideas of mathematics.

The thesis of this article is that classical mechanics and recent conceptual methods in abstract mathematics have a lot to do with each other. Many abstract ideas of pure geometry originally arose in the study of mechanics; for example, the cotangent bundle of a differentiable manifold appeared first as the phase space of Hamiltonian mechanics. Many cumbersome developments in the standard treatments of mechanics can be simplified and better understood when formulated with modern conceptual tools, as in the well-known case of the use of the "universal" definition of tensor products of vector spaces to simplify some of the notational excesses of tensor analysis as traditionally used in relativity theory. This article will develop a few more such cases. It has been stimulated by the pioneering work of Mackey [12], Smale [13], Abraham [1], Sternberg [14], and many others.

2. Quantities. Newton's Laws give differential equations for the trajectory of a mass point moving in Euclidean 3-space or along some surface or curve in that space. When there are several such points (say k of them) each moving in 3-space, they can be regarded as a single point in $3k$-space. Hence we consider Newton's Laws for a point moving in *any* "configuration space" M. On M we

are to deal with various measurable physical "quantities" such as the distance of a point from the origin, or the potential energy of a point of given mass, or the x and y coordinates of a point in the plane, or the latitude and longitude of a point on an equatorial belt U of a sphere S^2. Each such quantity is a continuous real-valued function g defined on all the points of M or of some "piece" U of M. These quantities, taken together, have a number of formal properties: To define a quantity g on the union $U' \cup U''$ of two pieces, it suffices to define quantities g' on U' and g'' on U'', so that they agree on the intersection $U' \cap U''$; if g and h are quantities defined on U, so is any smooth function $f(g, h)$ of g and h. These properties (the list is not complete) which describe quantities on configuration space have a direct mathematical interpretation. Take configuration space to be a differentiable manifold (of class C^∞), the pieces U of the space M to be open sets, and the quantities defined on U to be those continuous functions $g: U \to \boldsymbol{R}$ to the real numbers $\boldsymbol{R}$ which are smooth functions (of class C^∞). We recall that such a manifold can be defined as a topological space together with a suitable function S which assigns to each open set U of the space the set $S(U)$ of those continuous functions $g: U \to R$ which are the smooth functions on U. The axioms on S include those suggested above; for example, the description of a quantity g on $U' \cup U''$ in terms of its restrictions g' and g'' to U' and U'' is the essential element in the assertion that S is a "sheaf". Using S, one also describes the smooth maps of M to another manifold N: A function $f: M \to N$ is continuous if the inverse image $f^{-1}V$ of any open set V of N is open in M; it is smooth if for each smooth $h: V \to \boldsymbol{R}$ the composite hf is smooth (i.e., $hf \in S(f^{-1}V)$).

The axioms for a manifold M also describe the dimension n of M in terms of the existence of coordinates. Early in mechanics it was noted that problems are often best treated not with the original (rectangular) coordinates $x^1, \cdots, x^n$, but in terms of other "curvilinear" coordinates, say $q^1, \cdots, q^n$. These appear in the axiom on M which requires for each point $a \in M$ a neighborhood U (a "coordinate neighborhood") and n smooth functions, $q^1, \cdots, q^n: U \to \boldsymbol{R}$ such that the correspondence $p \mapsto q^1(p), \cdots, q^n(p)$ is a diffeomorphism (one-one, onto, and smooth in both directions) of U to an open set in $\boldsymbol{R}^n$. Here the smooth functions f on $\boldsymbol{R}^n$ are to be exactly the functions $f(x^1, \cdots, x^n)$ of class C^∞ (continuous derivatives of all orders). Hence the axiom states that the smooth functions g on a coordinate neighborhood U are precisely the C^∞ functions $g = f(q^1, \cdots, q^n)$ of the n coordinates. Many basic properties of manifolds and mechanical systems can be understood and expressed more clearly when they are formulated in a fashion independent of any particular choice of coordinates. Thus at the very beginning we speak of a quantity g or of a map $g: U \to \boldsymbol{R}$, not of a smooth function $g(q^1, \cdots, q^n)$ of some coordinates.

This review points up the fact that the physicists' use of "quantities" anticipates the mathematicians' use of smooth function, is a starting point for the currently active subject of sheaf theory **[15]**, and matches the modern emphasis on coordinate-free presentation of geometry. Often a book on physics will contrast a mathematical presentation of an idea (using coordinates) with a physical

presentation (not using coordinates); in such cases the contrast is really between two mathematical presentations, one with coordinates and one coordinate-free.

3. Velocity and tangent vectors. With Newton's Laws we may write down the usual equations of motion for our point of mass m moving on the configuration space M. These equations involve both velocity and acceleration, and so are second order differential equations; initial conditions will be given by specifying the initial position (by coordinates $q^1, \cdots, q^n$) and initial velocity (by the corresponding coordinates $v^1, \cdots, v^n$) (the coordinate v^i is that often written as $\dot{q}^i$, where the dot suggests "time derivative"). Taken all together, these $2n$ coordinates are not coordinates on M, but coordinates on another smooth manifold of twice the dimension, the *tangent bundle* $T_\bullet M$ of M. This bundle may be described in physical terms as the manifold of all positions-and-velocities on M, or in pictorial terms as the manifold of all tangent arrows attached to points a of M, or more formally as the manifold $T_\bullet M$ whose points are all ordered pairs (a, u) where a is a point of M and u a tangent vector at that point. To complete its description as a manifold, one then specifies that the smooth functions are (locally) the smooth functions of the $2n$ coordinates $q^1, \cdots, q^n, v^1, \cdots, v^n$, taken in a coordinate neighborhood of $T_\bullet M$ which is composed of a coordinate neighborhood U on M together with *all* tangent vectors at all of its points. With this description of the tangent bundle $T_\bullet M$ we note that the function indicated by $(a, u) \mapsto a$ is a smooth map $\pi_\bullet : T_\bullet M \to M$, called the *projection* of the tangent bundle. Under this projection the inverse image of each point $a \in M$ is the tangent space $T_a M$, which consists of all vectors tangent to M at a. This is not only a submanifold, but a linear vector space, and the full linear group (of nonsingular $n \times n$ matrices) acts on this space. These data, present and suitably interrelated in the tangent bundle, are essential elements in the recent treatment of many geometric ideas by vector bundles, K-theory, fiber bundles, and fiber spaces (see Husemoller [7]): They were really discovered and used a long time ago in mechanics.

4. Tangent and cotangent bundles. But we are ahead of our story; to complete the description of the tangent bundle we need an invariant description of the tangent spaces $T_a M$. Sometimes tangent vectors (or vector fields) are described as operators on smooth functions; we want a more symmetric treatment. Recall [9] that every finite dimensional vector space W determines in conceptual fashion another vector space of the same dimension, the *dual space* W^* consisting of all linear functions $W \to \boldsymbol{R}$. If $W = T_a M$ is a tangent space, this dual is called the *cotangent space* $T^a M$ to the manifold M at a. We describe them together: Every path determines a tangent vector, every quantity a cotangent vector. In detail, let $g: M \to \boldsymbol{R}$ be a quantity defined on M (or at least in some neighborhood of the chosen point $a \in M$) and let h be a smooth *path* on M passing through the point a. If we use the coordinate t (for "time") on the real numbers (a manifold) $\boldsymbol{R}$, this path can be represented as a smooth map $h: \boldsymbol{R} \to M$ send-

ing the origin $t=0$ to the chosen point $a \in M$. Thus we have two smooth maps

$$\boldsymbol{R} \xrightarrow{h} M \xrightarrow{g} \boldsymbol{R};$$

their composite, differentiated at $t=0$, yields a real number

$$\langle g, h\rangle_a = \left[\frac{d}{dt}(g \circ h)\right]_{t=0}.$$

If h' is a second path through a we say that h is *tangent* to h' at a (in symbols, $h \equiv_a h'$) if and only if $\langle g, h\rangle_a = \langle g, h'\rangle_a$ for all quantities g. The *tangent vector* $\tau_a h$ to h at a is then defined to be just the equivalence class consisting of all h' with $h \equiv_a h'$. Dually, two quantities g and g' are *cotangent* at a (in symbols $g \equiv_a g'$) if and only if $\langle g, h\rangle_a = \langle g', h\rangle_a$ for all smooth paths h through a; then the differential (cotangent vector) $d_a g$ to g at a is just the equivalence class of all these g'. Let $T^a M$ be the set of all cotangent vectors at a. Now all quantities $f, g, \cdots$ defined near a form a real vector space under the operations of addition and multiplication by real scalars: These operations, transferred to equivalence classes $d_a f$, $d_a g$, make $T^a M$ a real vector space. Moreover, the formula above for $\langle g, h\rangle_a$ defines a pairing (or function of two variables) which assigns to a cotangent vector $d_a g$ and a tangent vector $\tau_a h$ a real number,

$$(d_a g, \tau_a h) \mapsto \langle d_a g, \tau_a h\rangle = \langle g, h\rangle_a.$$

Addition of tangent vectors is now defined so as to make this function bilinear. This makes the set $T_a M$ of all tangent vectors to M at a into a real vector space, and the pairing makes $T^a M$ its dual space; indeed each $d_a g$ is the linear function $\langle d_a g, -\rangle : T_a M \to \boldsymbol{R}$.

We emphasize that this conceptual description of tangent and cotangent vectors does match the intuitive picture of these vectors: A tangent vector to a path at a is an arrow measuring the velocity along the path; a cotangent vector is a set of level lines (contour lines straightened up) given by a quantity g defined near a, and the pairing of the cotangent vector with the tangent vector gives the rate of change of the level-lined quantity along the path.

This invariant description also agrees with the description of tangent vectors in terms of local coordinates $q^1, \cdots, q^n$ on M: The pairing is given by the familiar formula

$$\langle d_a g, \tau_a h\rangle = \sum_{i=1}^{n} \left(\frac{\partial g}{\partial q^i}\right)_a \left(\frac{dq^i}{dt}\right)_0, \qquad \frac{d}{dt} \quad \text{on} \quad \boldsymbol{R},$$

which gives the derivative of a composite function. In writing this formula we are dealing with n quantities $q^i : M \to \boldsymbol{R}$, the coordinates on M, and a fixed map $h : \boldsymbol{R} \to M$. By this map h the quantities q^i become quantities $q^i h : \boldsymbol{R} \to \boldsymbol{R}$ on the 1-dimensional manifold $\boldsymbol{R}$, and $(dq^i/dt)_0$ in the formula refers to the derivatives of those latter composite quantities with respect to the coordinate t on R. Thus the tangent vector $\tau_a h$ has the n coordinates

$$v^1 = \left(\frac{dq^1h}{dt}\right)_0, \cdots, v^n = \left(\frac{dq^nh}{dt}\right)_0,$$

the tangent vector space T_aM is n-dimensional, and the whole tangent bundle $T_\bullet M$ can be made a differentiable manifold in exactly one way so that the $2n$ local coordinates are $q^1, \cdots, q^n, v^1, \cdots, v^n$. At the same time in the cotangent vector space T^aM each vector d_af has the n coordinates

$$p_1 = \left(\frac{\partial g}{\partial q^1}\right)_a, \cdots, p_n = \left(\frac{\partial g}{\partial q^n}\right)_a,$$

and T^aM is n-dimensional. The cotangent bundle $T^\bullet M$ has points all pairs (a, d_af), and can be made a differentiable manifold in exactly one way so that the $2n$ quantities $q^1, \cdots, q^n, p_1, \cdots, p_n$ become local coordinates.

There is an evident function which sends every point (a, d_af) on the cotangent bundle to the original point a on configuration space. This is a smooth map $\pi^\bullet$, called the *projection* of the cotangent bundle, displayed as

$$\pi^\bullet : T^\bullet M \to M \text{ by } (a, d_af) \mapsto a.$$

In coordinates, the point $(q^1, \cdots, q^n, p_1, \cdots, p_n)$ projects to $(q^1, \cdots, q^n)$.

5. Differentials and differential forms. If g is a quantity defined on an open set U of M, it determines at each point a of U a cotangent vector d_ag. This determination is really a function from U to the cotangent bundle. We call this function the *differential* of g: It is a smooth map

$$dg: U \to T^\bullet M \quad \text{by} \quad a \mapsto (a, d_ag).$$

This map is a *cross section* of the cotangent bundle in the sense that the composite $\pi^\bullet \circ dg: U \to U$ is the identity. This cross section dg is just the usual "differential" of the function g.

In general *any* cross section θ of the cotangent bundle $T^\bullet M$ over U is called a *1-form* on U. In case U is a coordinate neighborhood with coordinates $q^1, \cdots, q^n$, these 1-forms θ are precisely the things which can be written as expressions

$$\theta = f_1dq^1 + \cdots + f_ndq^n$$

in terms of the 1-forms dq^i (the differentials of the coordinates) and arbitrary smooth functions f_i on U. In this expression the formal operations of addition (of two 1-forms) and multiplication (of a 1-form by a quantity) have the natural invariant description: Since cotangent vectors at a can be added, we can add two 1-forms θ, ψ by adding their values $\theta a + \psi a$ in each T^aM; since cotangent vectors can be multiplied by scalars, we can multiply θ by a quantity f by the rule $(f\theta)a = (fa)(\theta a)$. Under these operations the 1-forms on U constitute a module over the ring $S(U)$ of quantities on U. Observe by the way that this (classical? modern?) concept of module dominates many recent developments in algebra [**9**] and homological algebra [**10**].

The systematic use of cotangent bundles, 1-forms, and the like is a major tool in geometry (see, for examples, Flanders [4]). In a moment we shall see that the cotangent bundle is simply the physicists' phase space. First we pause to consider the dual concepts. A cross section X of the tangent bundle $T_{\bullet}M$ over U is called a *vector field* on U because it assigns, in a smooth way, a tangent vector X_a at a to each $a \in U$. Moreover, a vector field X operates on quantities, assigning to each quantity g on U the quantity Xg defined at each point a by $(Xg)(a) = \langle d_a g, X_a \rangle$; this is called the *Lie derivative* of g along the field X. For example, in a coordinate neighborhood we can take at each point a the path $h_{1,a}$ "along the q^1 axis"; this is the path specified in coordinates as $h_{1,a}(t) = (q^1 a + t, q^2 a, \cdots, q^n a)$. Then $a \mapsto \tau_a h_{1a}$ is a vector field, the field of "unit" tangent vectors along the q^1 axis. This vector field is usually written as $X_1 = \partial/\partial q^1$, because the coordinate formula above for the basic pairing shows that its action on a function g is given by $X_1 g = \partial g/\partial q^1$. In the coordinate neighborhood every vector field X can be written uniquely in the form

$$X = f_1 \frac{\partial}{\partial q^1} + \cdots + f_n \frac{\partial}{\partial q^n},$$

where the f_i are quantities in U. In this formula, the addition and scalar multiplication by quantities f are defined for vector fields just as for forms: Under these operations the vector fields on U form a module over the ring $S(U)$ of quantities on U.

6. Lagrange's equations. We return to mechanics, where Newton's Law gives the equations of motion for a point in configuration space in the familiar form

$$m_i(d^2x^i/dt^2) = F_i, \qquad i = 1, \cdots, n,$$

where the $x^1, \cdots, x^n$ are rectangular coordinates and F_i is the ith component of the force. These equations may be rewritten in more invariant form using the quantities which represent potential and kinetic energy. The *kinetic energy* T, in the familiar rectangular coordinates, has an expression such as

$$T = (1/2)\, m_1(v^1)^2 + \cdots + (1/2)\, m_n(v^n)^2 = (1/2)\, m_1(\dot{q}^1)^2 + \cdots + (1/2)\, m_n(\dot{q}^n)^2.$$

In other coordinates there will be different formulas, but *any* such T depends on the coordinates in the tangent bundle, so T is a quantity $T: T_{\bullet}M \to R$ on that bundle. (The letter T is overused here, thanks to separate traditions, in which plain T is a quantity (kinetic energy) and dotted T a functor, the tangent bundle functor.) In a conservative system, the ith component F_i of the force on the particle can be expressed as the quantity

$$F_i = -\frac{\partial V}{\partial x^i},$$

where V is the potential energy. This V, originally a quantity $V: M \to \mathbf{R}$ on the

configuration space M, is also a quantity on the tangent bundle $T_\bullet M$, namely the composite function $T_\bullet M \to M \to \boldsymbol{R}$. Therefore the difference

$$L = T - V: T_\bullet M \to \boldsymbol{R}$$

is also a quantity on $T_\bullet M$. It is called the *Lagrangian.* Now in the rectangular coordinates above,

$$m_i \frac{d^2x^i}{dt^2} = \frac{d}{dt}\left(\frac{\partial T}{\partial \dot{x}^i}\right) = \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}^i}\right), \qquad F_i = -\frac{\partial V}{\partial x^i} = \frac{\partial L}{\partial x^i},$$

so the familiar equations of motion can be written as

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}^i}\right) - \frac{\partial L}{\partial x^i} = 0, \qquad i = 1, \cdots, n.$$

Now transform to *any* other local coordinates $q^1, \cdots, q^n$; a calculation shows that these equations again take the form

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}\right) - \frac{\partial L}{\partial q^i} = 0, \qquad i = 1, \cdots, n.$$

The *Lagrange equations* have the *same* form in *any* local coordinates; we omit the classical explanation: These are the equations for the extremal curves (trajectories) which minimize a suitable integral.

7. The Legendre transformation. Now we shift from the tangent bundle to the cotangent bundle. This shift to "phase space" is accomplished by a "Legendre transformation" which uses the quantity L to carry tangent spaces to cotangent spaces. This transformation, which sometimes appears adventitious when presented via coordinates, can be described invariantly, following an observant suggestion of Sternberg [**14**]; we start with the case of one tangent space (at one point $a \in M$) and get the following situation:

THEOREM. *If W is a real vector space, regarded as a differentiable manifold, then at each point $u \in W$ the tangent and cotangent spaces have canonical isomorphisms $T_uW \cong W$, $T^uW \cong W^*$. Any quantity L on W determines by $u \mapsto d_uL$ a smooth mapping $\mathfrak{L}_L: W \to W^*$, called the Legendre transformation for L. In particular, if L is a positive definite quadratic form on W, making W an inner product space, then $\mathfrak{L}_L$ is the standard natural isomorphism which identifies the inner product space W with its dual. If $\mathfrak{L}_L$ is invertible, then the inverse $(\mathfrak{L}_L)^{-1}$ is also a Legendre transformation for the quantity H on W^* which is defined in terms of (linear) coordinates $v^1, \cdots, v^n$ on W and the dual coordinates $p_1, \cdots, p_n$ on W^* as*

$$H = v^1p_1 + \cdots + v^np_n - L. \tag{1}$$

Proof. Each vector $w \in W$ determines a path $h_w: \boldsymbol{R} \to W$ at $u \in W$ by $t \mapsto u + tw$; this is just the "radial" path from the point u along the vector w. The correspondence $w \mapsto \tau_u h_w$ gives the desired isomorphism $W \cong T_uW$; its dual is $T^uW \cong W^*$.

The mapping $\mathcal{L}_L: W \to W^*$ comes from the very definition of the differential; if we describe differentials by coordinates as above, the ith coordinate of $\mathcal{L}_L u$, in symbols $p_i(\mathcal{L}_L u)$, is just the value $(\partial L/\partial V^i)_u$ of the ith partial derivative of L at this point u. Thus we can write (with $v^i = \dot{q}^i$)

$$p_i \circ \mathcal{L}_L = \partial L/\partial v^i = \partial L/\partial \dot{q}^i, \qquad i = 1, \cdots, n. \tag{2}$$

These equalities (between quantities $W \to \boldsymbol{R}$) describe the map $\mathcal{L}_L: W \to W^*$ completely in terms of its composites with coordinates on W^*. These equations are often written just as $p_i = \partial L/\partial \dot{q}^i$. This coordinate description of $\mathcal{L}$ is to be understood as an equation between quantities on W; we emphasize that it is just a coordinate version of the simple conceptual assignment $u \mapsto d_u L \in W^*$; in words, send each point of u to the value at u of the differential dL, where the space $T^u W$ of differentials is identified with W^* by the canonical isomorphism we have described.

Next suppose that L is a quadratic form. This is the important case for our purposes, in which W is the tangent space $T_a M$ to a configuration space M at some point a, while $L = T - V$ is the classical Lagrangian. Restricted to one tangent space $T_a M$ this L is (up to a constant) just the familiar positive definite form $\frac{1}{2}\sum m_i v_i^2$ for kinetic energy. This also provides an interpretation for the coordinates p_i, which we introduced just as the coordinates on W^* dual to the given v^i; with this L, $p_i = \partial L/\partial v^i$ is just $p_i = m_i v^i$, the ith-component of the *momentum*. For this reason the $p_1, \cdots, p_n$ are called the *momentum coordinates* in W^*.

Now the statement that the function L is a "quadratic form" is usually explained by saying that L has a quadratic expression in some coordinate system, but it can be also explained in invariant fashion: L is a function $L: W \to \boldsymbol{R}$ with $L(bu) = b^2 L(u)$ for any scalar b and such that the expression

$$(u, u') = (1/2)\left[L(u + u') - L(u) - L(u')\right]$$

is a bilinear function—that is, a function linear in each of the vectors u and $u' \in W$. (In other words, a function L is quadratic when its deviation $L(u+u') - L(u) - L(u')$ from additivity is bilinear.) Now the formula above gives a real inner product (u, u') for any two vectors of W; this inner product is symmetric in u and u' by definition, and is positive definite by the assumption on L. Hence W with L is an inner product space (a Euclidean vector space = a finite dimensional Hilbert space) and L is given in terms of the inner product as $L(u) = (u, u)$.

We return to the proof of the theorem. Along the standard path h_w which we introduced above we have

$$L(u + tw) = (u + tw, u + tw) = (u, u) + 2t(u, w) + t^2(w, w),$$

since the inner product is bilinear. Differentiating at $t = 0$ gives

$$\langle d_u L, w\rangle = 2(u, w).$$

In other words (up to the factor 2), $d_u L$ is that linear form on W given by the

function $(u, -)\colon W \to W$. This correspondence $u \mapsto d_u L$ is thus precisely the familiar map $u \mapsto (u, -)$ sending W to W^*, for W an inner product space.

The map $\mathcal{L}_L\colon W \to W^*$ will be invertible (locally) provided the familiar Jacobian matrix $\|\partial^2 L/\partial v^i \partial v^j\|$ is nonzero at every point of W. When it is invertible, the equation

$$H = \sum_{i=1}^{n} v^i p_i - L \qquad \text{on } W^* \text{ along } W \xleftarrow{\mathcal{L}_L^{-1}} W^*$$

does define a (smooth) quantity H on W^*. Specifically, in this equation the coordinates p_i on W^* are given as quantities on W^*, while the coordinates v^i and L, originally given as quantities on W, become quantities on W^* by composition with $\mathcal{L}_L^{-1}$. The equation is thus short for the more explicit equation

$$H = \sum_i (v^i \mathcal{L}_L^{-1}) p_i - L\mathcal{L}_L^{-1}$$

(and we emphasize this point because it explains a general method of reading an equation in quantities defined on different spaces in terms of quantities *on* one space *along* a given diagram $W \leftarrow W^*$ of arrows between those spaces). From this formula we may now calculate the Legendre transformation $\mathcal{L}_H\colon W^* \to W$. The coordinate formula for $\mathcal{L}$ (above) is $v^j \mathcal{L}_H = \partial H/\partial p_j$, with partial derivative calculated on W^*. By the usual rules for the derivative of a product

$$\frac{\partial H}{\partial p_j} = v^j + \sum_i \frac{\partial v^i}{\partial p_j} \cdot p_i - \sum_i \frac{\partial L}{\partial v^i} \frac{\partial v^i}{\partial p_j} = v^j;$$

the two sums cancel because $p_i \circ \mathcal{L}_L$ is $\partial L/\partial v^i$, by the definition of $\mathcal{L}_L$. This formula proves that $v^j \mathcal{L}_H \mathcal{L}_L = v^j$, so $\mathcal{L}_H \mathcal{L}_L = 1$ and $\mathcal{L}_H$ is indeed (locally) the inverse of $\mathcal{L}_L$. This completes the proof.

The formula for H may be stated in invariant terms, for any $y \in W^*$, as

$$Hy = \langle y, \mathcal{L}_L^{-1} y\rangle - L\mathcal{L}_L^{-1} y.$$

Here $\mathcal{L}_L^{-1} y \in W$ and $\langle , \rangle$ denotes the pairing $W^* \times W \to \boldsymbol{R}$ which makes W^* the dual of W. With coordinates $p_1, \cdots, p_n$ for y and dual coordinates $v^1, \cdots, v^n$ for $\mathcal{L}_L^{-1} y$, the value of this pairing is exactly the sum $\sum p_i v^i$ of the previous formula. The above calculation that $\mathcal{L}_H$ is $\mathcal{L}_L^{-1}$ may also be put in invariant form, but the translation to this form appears to be cumbersome.

The notation for this theorem has been chosen so as to fit the natural extension from one tangent space to the whole tangent bundle of any manifold M. Let L be a quantity on $T_\bullet M$, and identify each $T_u(T_a M)$ with $T_a M$ and $T^u(T_a M)$ with $(T_a M)^* = T^a M$, as above. A point of $T_\bullet M$ is then a pair (a, u) for $u \in T_a M$, and the quantity L determines, by restriction to each tangent space, a quantity $L| T_a M$ on that space. Then we define

$$\mathcal{L}_L\colon T_\bullet M \longrightarrow T^\bullet M, \qquad (a, u) \mapsto (a, d_u(L| T_a M)),$$

a smooth mapping of the tangent bundle $T_\bullet M$ to the cotangent bundle $T^\bullet M$. This mapping is sometimes called the *fiber derivative* of L (see Abraham [1]) because it arises by taking the differential of L restricted to each fiber. If we put in the projections of these two bundles, the diagram

$$\begin{tikzcd} T_\bullet M \arrow[r, "\mathfrak{L}_L"] \arrow[d, "\pi_\bullet"] & T^\bullet M \arrow[d, "\pi^\bullet"] \\ M \arrow[r, phantom, "="] & M \end{tikzcd}$$

commutes; one says that the Legendre transformation carries fibers (tangent spaces) to fibers (cotangent spaces). In coordinates, the formulas for $\mathfrak{L}_L$ are just those already given in (2), plus the equations $q^i\mathfrak{L}_L = q^i$ on the position coordinates. The case where L restricted to each fiber is a positive definite quadratic form is still applicable as above. Indeed, a manifold M with a smooth function assigning to each tangent space a quadratic form there is just a manifold with a *Riemann metric.* In other words, the kinetic energy in a classical mechanical system may be regarded as a Riemann metric on the configuration space.

8. Hamilton's Equations. This Legendre transformation $\mathfrak{L}$ is now applied to the trajectories of the given mechanical system. These trajectories are originally the paths $h: \boldsymbol{R} \to M$ on configuration space which satisfy the Lagrange equations. Now at each point $h(t)$ of M on such a path h the path itself determines a tangent vector $\tau_{h(t)}h$ to M, and the map $t \mapsto (h(t), \tau_{h(t)}h) \in T_\bullet M$ "lifts" the path h to a path $\tilde{h}: \boldsymbol{R} \to T_\bullet M$ on the tangent bundle; this is just the path giving position *and* velocity at any time $t \in \boldsymbol{R}$. The image under $\mathfrak{L}$ of these lifted trajectories will be the trajectories in $T^\bullet M$ of the mechanical system. A straightforward calculation from the Lagrange equations shows that these trajectories $\tilde{h}$ on $T^\bullet M$ are the solutions of the systems of equations

$$\frac{dq^i}{dt} = \frac{\partial H}{\partial p_i}, \qquad \frac{dp_i}{dt} = -\frac{\partial H}{\partial q^i}, \qquad i = 1, \cdots, n,$$

on $\boldsymbol{R}$, evaluated along $\tilde{h}: \boldsymbol{R} \to T^\bullet M$. Here "evaluated along $\tilde{h}$" means (as in a previous case) that each of the q^i, the p_i, and the partial derivatives of H, all originally quantities on $T^\bullet M$, are to be regarded as quantities on $\boldsymbol{R}$ (that is, as functions of t) by composition with $\tilde{h}$.

These $2n$ equations are *Hamilton's equations.* They describe the trajectories on *phase space* $T^\bullet M$ by a system of first order differential equations. These equations depend on the quantity H on $T^\bullet M$; this quantity is the *Hamiltonian* defined above in terms of the Lagrangian. The formulation of these equations also seems to depend on the fact that phase space is the manifold of "positions and momenta" (mathematically, that phase space is a cotangent bundle). This means that we can get these equations in this form for any system of $2n$ local coordinates on $T^\bullet M$, *provided* these coordinates are split into two sets: position

coordinates $q^1, \cdots, q^n$ and the corresponding momentum coordinates $p_1, \cdots, p_n$. Any position coordinates will do—provided that they are matched with the corresponding momentum coordinates. There is another mystery beyond the use of such pairs of coordinates. A system of first order differential equations on a manifold $T^\bullet M$ is usually given by a vector field X on $T^\bullet M$; the solutions of the equations may then be described as those paths on the manifold which thread through the vector field, with tangent vectors at each point b exactly that specified as Xb by the vector field at that point. But this system of equations is determined by the quantity H, and what a quantity H gives naturally is a covector field dH, not a vector field.

So we reconsider the properties of phase space. The formula $H = \sum q^i p_i - L$ used to define H suggests the expression

$$\theta = \sum p_i \, dq^i.$$

Now p_i and q^i are quantities on $T^\bullet M$, so each dq^i is a 1-form on $T^\bullet M$, and so is each scalar multiple $p_i dq^i$ and hence θ itself. Now if we make any change in the local coordinates q^i and the corresponding change in the momenta p_i, a calculation shows that this 1-form θ is unchanged. Indeed, this 1-form can be described in invariant fashion, using only the fact that $T^\bullet M$ is a cotangent bundle. As a 1-form on $T^\bullet M$, θ must be a cross section of the cotangent bundle of $T^\bullet M$; that is, a smooth map $T^\bullet M \to T^\bullet(T^\bullet M)$ which is a right inverse of the projection $\pi^{\bullet\bullet}: T^\bullet(T^\bullet M) \to T^\bullet M$. So take any point $b = (a, d_a g)$ on $T^\bullet M$ and define θ as the map

$$b = (a, d_a g) \mapsto ((a, d_a g), d_b(g\pi^\bullet)).$$

Here $\pi^\bullet: T^\bullet M \to M$ is the projection of the given tangent bundle, so $g\pi^\bullet$ is a quantity on $T^\bullet M$, and the expression on the right in this formula is a point of $T^\bullet(T^\bullet M)$ with first component $(a, d_a g)$ the given point b. Thus this formula does define a 1-form θ'. Moreover, $\theta' = \theta$: For $\theta = \sum p_i dq^i: T^\bullet M \to T^\bullet(T^\bullet M)$ is a linear combination of the 1-forms dq^i, so is given by the formula (definition of the linear operations $\sum_i$)

$$\theta b = (b, \sum p_i(b) d_b q^i).$$

Now we can write $b = (a, d_a g) = (a, \sum k_i \, d_a q^i)$ with n scalars k_i; these scalars being exactly the p_i coordinates $k_i = p_i \, b$. Therefore

$$\begin{aligned} \theta(b) = \theta(a, d_a g) &= (b, \sum k_i \, d_b q^i) \\ &= (b, \sum k_i \, d_b(q^i \pi^\bullet)) \\ &= \theta'(a, \sum k_i \, d_a q^i) = \theta'(a, d_a g), \end{aligned}$$

so $\theta = \theta'$, as claimed. On *every* cotangent bundle there is an invariant 1-form θ.

9. Forms and symplectic manifolds. On any manifold S we now consider 1-forms θ and 2-forms ω. In terms of local coordinates $x^1, x^2, \cdots$ these are ex-

pressions (in quantities f_i, g_i) like

$$\theta = f_1 dx^1 + f_2 dx^2 + \cdots, \quad \omega = g_{12} dx^1 \wedge dx^2 + g_{13} dx^1 \wedge dx^3 + \cdots;$$

and any 1-form θ determines a 2-form $d\theta$ by formal application of the differential d as

$$d\theta = \left(\frac{\partial f_2}{\partial x^1} - \frac{\partial f_1}{\partial x^2}\right) dx^1 \wedge dx^2 + \cdots.$$

The conceptual presentation uses the exterior algebra of the cotangent spaces $W^* = T^b S$ of S (see, for example [9]). First, the exterior square $W^* \wedge W^*$ is the vector space generated by all formal products $e \wedge e'$ of the vectors e, $e' \in W^*$, where the exterior product $\wedge$ is required to be bilinear and alternating ($e \wedge e' = -e' \wedge e$, hence $e \wedge e = 0$). If W has a basis of n vectors e, then $W^* \wedge W^*$ has a basis of $n(n-1)/2$ vectors $e^i \wedge e^j$ for $i < j$. Over the manifold S we can construct the exterior bundle $T^\bullet S \wedge T^\bullet S$ whose fiber over each point b is the vector space $T^b S \wedge T^b S$. A 2-form ω on S is then defined to be a smooth cross section of this bundle $T^\bullet S \wedge T^\bullet S \to S$. Since the dx^i for local coordinates x^i form a module basis of the 1-forms, the exterior algebra above shows that the $dx^i \wedge dx^j$ for $i < j$ form a module basis for the 2-forms. This means exactly that any 2-form can be written uniquely in the form $\sum g_{ij} dx^i \wedge dx^j$ displayed above.

In particular, on any cotangent bundle $S = T^\bullet M$, our invariant 1-form $\theta = \sum p_i \, dq^i$ yields by exterior differentiation a 2-form $\omega = d\theta$ with

$$\omega = \sum_{i=1}^{n} dp_i \wedge dq^i.$$

It is a closed 2-form ($d\omega = 0$ because $d\omega = dd\theta$) and is of maximal rank. The latter means that its exterior power $\omega^n = \omega \wedge \cdots \wedge \omega$ to n factors is nonzero; indeed $\omega^n = \pm k dp_1 \wedge \cdots \wedge dp_n \wedge dq^1 \wedge \cdots \wedge dq^n$ is (up to a factor k) just the familiar "element of $2n$-dimensional volume" on $T^\bullet M$.

Now an even dimensional manifold S equipped with a specified 2-form ω, with $d\omega = 0$, of maximal rank is called a *symplectic manifold.* We have just seen that every cotangent bundle (every phase space $T^\bullet M$) is a symplectic manifold. One reason for naming such manifolds is precisely that the Hamilton equation above can be written down directly for any quantity H on *any* symplectic manifold. This can be seen by choices of coordinates: Given any 2-form ω, with $d\omega = 0$ and of maximum rank, a theorem due to Darboux states that in some neighborhood of each point one can successively choose coordinates $q^1, \cdots, q^n$ and $p_1, \cdots, p_n$ so that ω takes on the standard form $\omega = \sum dp_i dq^i$. With these coordinates we can write down Hamilton's equations. Calculation then shows that a different choice of coordinates with the same properties yields the same trajectories for Hamilton's equation.

There is a better and more conceptual proof which uses the 2-form ω directly. If W^* is the dual of the vector space W, then each $e \in W^*$ is a linear function

$W \to \boldsymbol{R}$. Because of this, we shall show that each element $t \in W^* \wedge W^*$ yields a linear function

$$t^\flat: W \longrightarrow W^*.$$

Since any t is a linear combination of vectors $e \wedge e'$ for e and $e' \in W^*$, it will suffice to take $t = e \wedge e'$ (and use the corresponding sums of functions $t^\flat$). Thus we define $(e \wedge e')^\flat$ on each $w \in W$ to be an element of W^*; that is, to be the linear function of $w' \in W^*$ given by the formula

$$[(e \wedge e')^\flat w](w') = e(w)e'(w'),$$

where $e(w)$ and $e'(w')$ are real numbers, since e, $e': W \to \boldsymbol{R}$. Now consider the 2-form ω on S; at each point $b \in S$, $\omega_b \in T^b S \wedge T^b S$ gives a map $T_b S \to T^b S$ of tangent space to cotangent space. Put together at all points, this is a smooth map

$$\begin{tikzcd} T_\cdot S \arrow[r, "\omega^\flat"] \arrow[d] & T^\cdot S \arrow[d, shift right] \\ S \arrow[r, phantom, "="] & S \arrow[u, shift right, "dH"'] \end{tikzcd}$$

of the tangent bundle to the cotangent bundle. When ω is of maximal rank on S, this map is one-one and onto, so its inverse is a smooth map of the cotangent bundle back to the tangent bundle. In particular, given any 1-form on S, such as the differential dH of the Hamiltonian, this map gives a vector field $(\omega^\flat)^{-1} \circ dH$ on S. This is the vector field which is used in Hamilton's equations; therefore these equations arise directly from the quantity H and the basic 2-form ω on the symplectic manifold S.

10. Canonical transformations and generating functions. Specific problems in mechanics are often treated by picking coordinates convenient to the problem in hand; in the symplectic formulation of mechanics, this means picking on S new local coordinates, conventionally written $Q^1, \cdots, Q^n, P_1, \cdots, P_n$, in such a way that the 2-form ω is still expressed as $\omega = \sum dP_i \wedge dQ^i$. Classically, this is called a *canonical transformation* from the canonical coordinates p_i, q^i to the new canonical coordinates P_i, Q^i. These new coordinates need no longer have a direct physical interpretation by momentum and position.

Certain "generating functions" F are classically used to give such canonical transformations. The traditional presentation ([6], [16]) goes about as follows. Take a smooth "generating function" of $2n$ coordinates $q^1, \cdots, q^n, P_1, \cdots, P_n$, and assume the determinant

$$\det \left\| \frac{\partial^2 F}{\partial q^i \partial P_j} \right\| \neq 0$$

everywhere. Now the definitions

$$q^i = q^i, \qquad p_i = \partial F / \partial q^i, \qquad i = 1, \cdots, n,$$

give $2n$ functions which we regard as coordinates, while the parallel definitions

$$Q^i = \partial F/\partial P_i, \qquad P_i = P_i, \qquad i = , \cdots, n$$

again give $2n$ coordinates. The determinant condition implies that we can solve for the $Q^1, \cdots, Q^n, P_1, \cdots, P_n$ as functions of the $q^1, \cdots, q^n, p_1, \cdots, p_n$ or conversely. A calculation (to be done below) shows that this transformation carries the form $\sum dP_i \wedge dQ^i$ to the form $\sum dp_i \wedge dq^i$. Thus F has "generated" a transformation from one set of canonical coordinates to another.

This situation could also be described as follows. Let S be a manifold of even dimension $2n$ which is given as a product $S = S_1 \times S_2$ of two n-dimensional manifolds S_1 and S_2, and let F be a quantity on S. If α is any form on S (a 2-form, a 1-form, or a 0-form; that is, a quantity such as F), then the exterior derivative $d\alpha$ can be written in a natural way as a sum $d\alpha = d_1\alpha + d_2\alpha$, where d_1 is the differential along S_1 and d_2 that along S_2. In particular, d_1F is a 1-form on S, and $d_2(d_1F) = d_1(d_2F)$ is a 2-form on S. Our basic assumption now reads: F is a quantity on a manifold $S = S_1 \times S_2$ such that the 2-form $\omega = d_1d_2F$ has maximal rank. This assumption means exactly that S with the 2-form ω is a symplectic manifold. Indeed $d\omega = 0$ becomes

$$d\omega = d(d_1d_2\, F) = (d_1 + d_2)(d_1d_2F) = d_1^2d_2F + d_1d_2^2F = 0.$$

Choose any local coordinates $q^1, \cdots, q^n$ on S_1 and $P_1, \cdots, P_n$ on S_2. Then $q^1, \cdots, q^n, P_1, \cdots, P_n$ are $2n$ local coordinates on $S = S_1 \times S_2$, and

$$d_1F = \frac{\partial F}{\partial q^1}\, dq^1 + \cdots + \frac{\partial F}{\partial q^n}\, dq^n,$$

hence, applying d_2 in the evident way,

$$\omega = d_2d_1F = \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2 F}{\partial q^i \partial P_j}\, dq^i \wedge dP_j.$$

This is the formula by which the quantity F defines a 2-form. Calculating the exterior product $\omega \wedge \cdots \wedge \omega$, to n factors, the condition that it is nonzero becomes exactly the condition that the determinant above is nonzero at all points of S.

From the data $S = S_1 \times S_2$ and F we have thus defined a symplectic manifold $(S, \omega = d_1d_2F)$. In this manifold (and in these coordinates) the basic 2-form does *not* have the usual canonical form. But there are other local coordinates on S; for example, the determinant condition above implies that the quantities $q^1, \cdots, q^n, p_1, \cdots, p_n$ with $p_i = \partial F/\partial q^i$ are also local coordinates. In these coordinates we may calculate the 2-form $\omega_0 = \sum dp_i \wedge dq^i$ by the usual formula

$$\omega_0 = \sum_i dp_i \wedge dq^i = \sum_i \sum_j \frac{\partial^2 F}{\partial p_j \partial q^i} dp_j \wedge dq^i + \sum_i \sum_j \frac{\partial^2 F}{\partial q^j \partial q^i}\, dq^j \wedge dq^i.$$

The first double summation is $\omega = d_1d_2F$, while the second contains terms with

$dq^i \wedge dq^i = 0$, and terms with $dq^j \wedge dq^i = -dq^i \wedge dq^j$ and $i \neq j$, which cancel in pairs; therefore $\omega_0 = \omega$. In the coordinates q^i, p_i, ω becomes canonical. An exactly analogous calculation shows that ω is canonical in the coordinates P_i, $Q^i = \partial F / \partial P_i$. These two calculations together give the calculations previously suggested which show that the transformation from the coordinates q^i, p_i to Q^i, P_i takes the canonical 2-form to itself. But the calculation is best understood by going through the form ω expressed in terms of the intermediate (or original) coordinates $q^1, \cdots, q^n, P_1, \cdots, P_n$.

The same intermediate step can be exhibited in other classical cases (see [6], [16]) of generating functions for canonical transformations. The presentation of these ideas in classical books is sometimes impoverished because the notion of a differential form (other than a 1-form) is wholly avoided. The general methods of E. Cartan for invariant integrals use the calculus of differential forms extensively and go far beyond what we have presented here.

We have considered only changes of coordinates. Dually, there are symplectic transformations between two symplectic manifolds (S, ω) and (S', ω') of the same dimension. These are the smooth maps $f: S \to S'$ which carry the form ω' on S' back to the form ω on S (in the natural sense in which forms go backwards).

11. Hamilton-Jacobi Equations. Many problems of classical mechanics can be solved by first obtaining solutions of a certain first order partial differential equation, the Hamilton-Jacobi equation. We sketch the situation very briefly. Given the Hamiltonian $H: T^* M \to \mathbf{R}$, each quantity S on M determines a 1-form $dS: M \to T^* M$ which is a cross section of the cotangent bundle. The composite of these two functions is a quantity on M; setting it equal to 0 gives an equation

$$H \circ dS = 0 \qquad \text{on } M$$

called the *Hamilton-Jacobi* partial differential equation. If we write the Hamiltonian H as a function $H(q^1, \cdots, q^n, p_1, \cdots, p_n)$ of local coordinates and the differential $dS: M \to T^* M$ as the map $a \mapsto (a, d_a S)$, where a has the coordinates $q^1, \cdots, q^n$, this equation appears as

$$H\left(q^1, \cdots, q^n, \frac{\partial S}{\partial q^1}, \cdots, \frac{\partial S}{\partial q^n}\right) = 0;$$

it is a partial differential equation of the first order in S.

On the other hand, for any path $h: \mathbf{R} \to M$, giving the quantity S on M allows us to lift this path to $\tilde{h} = dS \circ h: \mathbf{R} \to T^* M$, a path on the cotangent bundle. The role of the Hamilton-Jacobi partial differential equations may then be formulated in the following theorem (which I take from some lecture notes of George Mackey, who treated the more general case of a "time dependent" Hamiltonian):

THEOREM. *Let S be a solution of the Hamilton-Jacobi partial differential equations for a Hamiltonian quantity H on a cotangent bundle $T^{*}M$. Then, if a path h on M satisfies*

$$\frac{dq^i}{dt} = \frac{\partial H}{\partial p_i}, \qquad i = 1, \cdots, n \quad \textit{on } \mathbf{R} \textit{ along } h \textit{ and } \tilde{h},$$

(the first n Hamilton equations) it also satisfies the second n equations

$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial q^i}, \qquad i = 1, \cdots, n \textit{ on } \mathbf{R} \textit{ along } h \textit{ and } \tilde{h}.$$

Conversely, if S is a quantity on M which has this property for all h, then S satisfies the Hamilton-Jacobi partial differential equation.

There are many further connections with the theory of characteristics of first order partial differential equations. This beautiful theory—well presented, say, in [3] and [5], is normally treated analytically, with heavy use of coordinates. It should be possible and useful to have a coordinate-free and more geometrical presentation for this and similar "analytic" theories.

12. Summary. With these lines we have just begun the exploration of the relation between ideas arising in classical mechanics and conceptual methods of modern differential geometry. Much more could be said in differential geometry (see [2], [14]) or in the translation of the classical presentation of mechanics ([6], [16], [18], etc.) into geometrical language. Some portions of this task are carried out at an elementary level in [8] and [11], and at a more advanced level in [1], [12], [17]. We have not touched on the active research in the related topics of structural stability and qualitative dynamics. We have emphasized the idea that quantities are functions and that equations hold on a manifold along a diagram of functions. More use of categorical ideas may be indicated. There has been a suggestion (F. W. Lawvere) that "categorical dynamics" will replace the category of differentiable manifolds by other more flexible categories, axiomatically characterized so as to bring out the tangent bundle and infinitesimal constructions. Here, as throughout mathematics, conceptual methods should penetrate deeper to give clearer understanding.

References

1. Ralph Abraham, Foundations of Mechanics, Benjamin, New York, 1967.
2. R. L. Bishop and S. I. Goldberg, Tensor Analysis on Manifolds, Macmillan, New York, 1968.
3. Constantin Caratheodory, Calculus of Variations and Partial Differential Equations of the First Order (Translated from the German), Holden-Day, San Francisco, 1965.
4. Harley Flanders, Differential Forms, with Applications to the Physical Sciences, Academic Press, New York, 1963.
5. Philipp Frank and Richard von Mises, Die Differentialgleichungen der Mechanik und Physik, 2nd edition, Vieweg, Braunschweig, 1930 and 1935.
6. Herbert Goldstein, Classical Mechanics, Addison-Wesley, Cambridge, Mass., 1951.

7. Dale Husemoller, Fibre Bundles, McGraw-Hill, New York, 1966.

8. L. H. Loomis and S. Sternberg, Advanced Calculus, Addison-Wesley, Reading, Mass., 1968.

9. Saunders Mac Lane and Garrett Birkhoff, Algebra, Macmillan, New York, 1967.

10. Saunders Mac Lane, Homology, Springer, Heidelberg and New York, 1963.

11. ———, Geometrical Mechanics, Parts I and II. Lecture notes by Raphael Zahler, assisted by George Angwin, William Corvette, David Golber, Paul Palmquist, Reinhard Schultz, Martha Smith, and Stanley Weiss. Department of Mathematics, The University of Chicago, 1968.

12. George W. Mackey, The Mathematical Foundations of Quantum Mechanics, Benjamin, New York, 1963.

13. S. Smale, Differentiable dynamic systems, Bull. Amer. Math. Soc., 73 (1967) 747–817.

14. Schlomo Sternberg, Lectures on Differential Geometry, Prentice-Hall, Englewood Cliffs, N. J., 1964.

15. Richard G. Swan, The Theory of Sheaves, The University of Chicago Press, 1964.

16. E. T. Whittaker, A Treatise on the Analytical Dynamics of Particles and Rigid Bodies, 4th edition, Cambridge University Press, 1959.

17. R. Hermann, Differential Geometry and the Calculus of Variations, Academic Press, New York, 1968.

18. L. A. Pars, A Treatise on Analytical Dynamics, Wiley, New York, 1965.

ON SOME PROBABILISTIC ASPECTS OF CLASSICAL ANALYSIS

MARK KAC, Rockefeller University

1. When one looks upon the numerous connections between probability theory and analysis one notes that the interactions between the two fall into two categories.

On the one hand we have problems whose very formulation depends on concepts and terminology of probability theory, e.g., what is the probability that a power series

$$\sum_{0}^{\infty} \pm a_n z^n \tag{1.1}$$

with random signs has its circle of convergence as a natural boundary?

On the other hand we have a more subtle and less direct intervention of probability theory in the situation in which neither the formulation nor the solution of the problem has anything to do with probability.

Today I should like to illustrate this second category of interactions by discussing a number of points in the classical theory of the logarithmic potential.

2. I begin with a cursory review of the Wiener measure in the space of plane curves, or equivalently, of the theory of two-dimensional Brownian motion.

Let O be an origin chosen in the plane and denote by $\vec{r}(t)$, $0 \leqq t < \infty$, continuous plane curves originating at O.

There is a completely additive (Lebesgue) measure P in the space of these curves which has the following fundamental property.

Let $0 \leqq t_1 < t_2 < \cdots < t_n$ and let $\Omega_1, \Omega_2, \cdots$ be plane Borel sets, then

$$(2.1)\quad P\{\vec{r}(t_1) \in \Omega_1, \vec{r}(t_2) \in \Omega_2, \cdots, \vec{r}(t_n) \in \Omega_n\}$$
$$= \int_{\Omega} \cdots \int_{\Omega_n} p(0 \mid \vec{r}_1; t_1) p(\vec{r}_1 \mid \vec{r}_2; t_2 - t_1) \cdots p(\vec{r}_{n-1} \mid \vec{r}_n; t_n - t_{n-1}) d\vec{r}_1 \cdots d\vec{r}_n,$$

where the integration with respect to $\vec{r}_k$ is over Ω_k and

$$(2.2)\qquad p(\vec{\rho} \mid \vec{r}; t) = \frac{\exp[-\|\vec{r} - \vec{\rho}\|^2/2t]}{2\pi t}$$

($\|\vec{a}\|$ denotes the Euclidean length of $\vec{a}$). The integral with respect to the (Wiener) measure P will be denoted by E to conform to standard probabilistic terminology.

It should perhaps be noted that (2.1) (with p given by (2.2)) is taken from the *physical theory* of plane Brownian motion. Wiener's great achievement was to show that assigning the measure (2.1) to sets

$$(2.3)\qquad \{\vec{r}(t_1) \in \Omega_1, \cdots, \vec{r}(t_n) \in \Omega_n\}$$

of *continuous* curves can be *extended* to a vastly richer class of sets. There is, in fact, a complete analogy between what Wiener did to the space of all continuous curves and what Lebesgue did to the real line: when starting by assigning a measure to *intervals* (sets (2.3) are the analogues of intervals and were called "quasi-intervals" by Wiener) he extended it to a large class of sets of real numbers.

3. Let Ω be a bounded, closed plane region and let A be an open set in the plane. Denoting by V_Ω the indicator function of Ω (and similarly by V_A the indicator function of A) we consider the expression

$$(3.1)\quad \int_0^\infty e^{-st} E\left\{\exp\left[-u\int_0^t V_\Omega(\vec{y} + \vec{r}(\tau))d\tau\right] V_A(\vec{y} + \vec{r}(t))\right\} dt, \quad u > 0, s > 0.$$

The motivation for considering (3.1) comes from the fact that

$$\int_0^t V_\Omega(\vec{y} + \vec{r}(\tau))d\tau$$

is the time, up to t, spent by the Brownian particle (which starts from $\vec{y}$) in the set Ω and the joint probability

$$(3.2)\qquad P\left\{\int_0^t V_\Omega(\vec{y} + \vec{r}(\tau))d\tau < \alpha, V_A(\vec{y} + \vec{r}(t)) = 1\right\}$$

that this occupation time is less than α *and* that the path ends up at t in A, is clearly of interest.

Expression (3.1) is a double Laplace transform of (3.2) (with respect to α and t), and for technical reasons is easier to study than (3.2).

It is now quite easy to show that

$$\int_0^\infty e^{-st}E\left\{\exp\left[-u\int_0^t V_\Omega(\vec{y}+\vec{r}(\tau))d\tau\right]V_A(\vec{y}+\vec{r}(t))\right\}dt = \int_A F(\vec{r};u,s)d\vec{r}, \tag{3.3}$$

where $F(\vec{r};u,s)$ satisfies the integral equation

$$F(\vec{r};u,s) = \frac{1}{\pi}K_0(\sqrt{2s}\|\vec{y}-\vec{r}\|) - \frac{u}{\pi}\int_\Omega K_0(\sqrt{2s}\|\vec{\rho}-\vec{r}\|)F(\vec{\rho};u,s)d\vec{\rho}, \tag{3.4}$$

where K_0 is the familiar Bessel function which makes its appearance through the formula

$$\int_0^\infty e^{-st}\,\frac{\exp[-\|\vec{r}-\vec{y}\|^2/2t]}{2\pi t}\,dt = \frac{1}{\pi}K_0(\sqrt{2s}\|\vec{r}-\vec{y}\|). \tag{3.5}$$

Let me derive (3.4) formally so that you will see a little bit the kind of manipulations one has to perform.

We have, expanding the exponential,

$$\begin{aligned} &E\left\{\exp\left[-u\int_0^t V_\Omega(\vec{y}+\vec{r}(\tau))d\tau\right]V_A(\vec{y}+\vec{r}(t))\right\} \\ &\qquad = \sum_{k=0}^\infty \frac{(-u)^k}{k!}E\left\{\left(\int_0^t V_\Omega(\vec{y}+\vec{r}(\tau))d\tau\right)^k V_A(\vec{y}+\vec{r}(t))\right\}, \end{aligned} \tag{3.6}$$

and note that

$$\begin{aligned} &E\left\{\left(\int_0^t V_\Omega(\vec{y}+\vec{r}(\tau))d\tau\right)^k V_A(\vec{y}+\vec{r}(t))\right\} \\ &= k!\underset{0\le t_1\le t_2\le\cdots\le t_k\le t}{\int\cdots\int} P\{\vec{y}+\vec{r}(t_1)\in\Omega,\cdots,\vec{y}+\vec{r}(t_k)\in\Omega,\vec{y}+\vec{r}(t)\in A\}dt_1\cdots dt_k \\ &= k!\underset{0\le t_1\le t_2\le\cdots\le t_k\le t}{\int\cdots\int}\int_\Omega\cdots\int_\Omega\int_A p(\vec{y}\,|\,\vec{r}_1;t_1)p(\vec{r}_1\,|\,\vec{r}_2;t_2-t_1)\cdots \\ &\qquad\cdot p(\vec{r}_{k-1}\,|\,\vec{r}_k;t_k-t_{k-1})p(\vec{r}_k\,|\,\vec{r};t-t_k)d\vec{r}_1\cdots d\vec{r}_k\,d\vec{r}\,dt_1dt_2\cdots dt_k, \end{aligned} \tag{3.7}$$

where the basic formula (2.1) has been used.

Performing the integrations on t's first, and using the familiar convolution theorem for Laplace transforms we get

$$
\begin{aligned}
F(\vec{r};u,s) &= \frac{1}{\pi}K_0(\sqrt{2s}\|\vec{r}-\vec{y}\|) \\
&\quad - \frac{u}{\pi}\sum_{k=1}^{\infty}(-u)^{k-1}\int_\Omega\cdots\int_\Omega K_0(\sqrt{2s}\|\vec{r}_1-\vec{y}\|)\cdots \\
&\qquad K_0(\sqrt{2s}\|\vec{r}-\vec{r}_k\|)d\vec{r}_1\cdots d\vec{r}_k
\end{aligned}
\tag{3.8}
$$

and (3.4) follows almost at once.

There are, of course, questions of convergence. For example (3.8) holds only for sufficiently small u. But a simple argument based on analytic continuation ("permanence of formulas") establishes (3.4) for *all* positive u.

4. Using the familiar asymptotic formula

$$
K_0(x) = \log\frac{2}{x} - \gamma + o(1), \qquad x\to 0\ (\gamma\text{—the Euler's constant})
\tag{4.1}
$$

we have

$$
K_0(\sqrt{2s}\|\vec{r}-\vec{\rho}\|) = \log\sqrt{2/s} + \log\frac{1}{\|\vec{r}-\vec{\rho}\|} - \gamma + o(1),
\tag{4.2}
$$

where $o(1)$ is uniform in $\vec{r}$ and $\vec{\rho}$, provided both are restricted to a compact set.

We can rewrite now (3.4) in the form

$$
\begin{aligned}
F(\vec{r}) &= 1/\pi\{K_0(\sqrt{2s}\|\vec{y}-\vec{r}\|) - \log\sqrt{2/s}\} \\
&\quad - \frac{u}{\pi}\int_\Omega\{K_0(\sqrt{2s}\|\vec{r}-\vec{\rho}\|) - \log\sqrt{2/s}\}F(\vec{\rho})d\vec{\rho} \\
&\quad + \log\sqrt{2/s}\left\{1 - u\int_\Omega F(\vec{\rho})d\vec{\rho}\right\}.
\end{aligned}
\tag{4.3}
$$

I shall now ask you to believe me that one can prove that as $s\downarrow 0, F(\vec{r}) = F(\vec{r};u,s)$ remains bounded and since as $s\downarrow 0, F(\vec{r};u,s)$ is monotonically increasing, the existence of the limit

$$
\lim_{s\downarrow 0} F(\vec{r};u,s) = f(\vec{r};u)
\tag{4.4}
$$

follows. From (4.3) and (4.2) we conclude that

$$
\lim_{s\downarrow 0}\log\sqrt{2/s}\left\{1 - u\int_\Omega F(\vec{\rho};u,s)d\vec{\rho}\right\} = g_\Omega(\vec{y};u)
\tag{4.5}
$$

exists.

5. Let

$$
G(\vec{y};u,s) = s\int_0^\infty e^{-st}E\left\{\exp\left[-u\int_0^t V(\vec{y}+\vec{r}(\tau))d\tau\right]\right\}dt
\tag{5.1}
$$

and note first that

$$G(\vec{y}; u, s) = 1 - u\int_\Omega F(\vec{\rho}; u, s)d\vec{\rho}. \tag{5.2}$$

A formal derivation (which can, however, be rigorized) of (5.2) is as follows. By (3.3) we have

$$\int_\Omega F(\vec{\rho}; u, s)d\vec{\rho} = \int_0^\infty e^{-st}E\left\{\exp\left[-u\int_0^t V_\Omega(\vec{y} + \vec{r}(\tau))d\tau\right]V_\Omega(\vec{y} + \vec{r}(t))\right\}dt$$
$$= -\frac{1}{u}\int_0^\infty e^{-st}\frac{d}{dt}E\left\{\exp\left[-u\int_0^t V_\Omega(\vec{y} + \vec{r}(\tau))d\tau\right]\right\}dt,$$

and (5.2) follows by integration by parts if only

$$\lim_{t\to\infty} E\left\{\exp\left[-u\int_0^t V_\Omega(\vec{y} + \vec{r}(\tau))d\tau\right]\right\} = 0, \qquad u > 0. \tag{5.3}$$

The truth of (5.3) follows at once from the well-known and relatively elementary fact that in the plane

$$\lim_{t\to\infty}\int_0^t V_\Omega(\vec{y} + \vec{r}(\tau))d\tau = \infty. \tag{5.4}$$

It is also easy to show (the formal steps are essentially the same as in Section 3) that G satisfies the integral equation

$$G(\vec{y}; u, s) = 1 - \frac{u}{\pi}\int_\Omega K_0(\sqrt{2s}\|\vec{\rho} - \vec{y}\|)G(\vec{\rho}; u, s)d\vec{\rho} \tag{5.5}$$

which we rewrite in the form

$$\begin{aligned} G(\vec{y}; u, s) &= 1 - \frac{u}{\pi}\int_\Omega (\log\sqrt{2/s})G(\vec{\rho}; u, s)d\vec{\rho} \\ &\quad - \frac{u}{\pi}\int_\Omega (K_0(\sqrt{2s}\|\vec{\rho} - \vec{y}\|) - \log\sqrt{2/s})\, G(\vec{\rho}; u, s)d\vec{\rho}. \end{aligned} \tag{5.6}$$

Multiplying both sides of (5.6) by $\log(\sqrt{2/s})$ and recalling that by (4.5) the limit

$$\lim_{s\downarrow 0}\log\sqrt{2/s}\, G(\vec{y}; u, s) = \lim_{s\downarrow 0}\log\sqrt{2/s}\left(1 - u\int_\Omega F(\vec{r}; u, s)d\vec{r}\right) = g_\Omega(\vec{y}; u) \tag{5.7}$$

exists, we obtain

$$g_\Omega(\vec{y}; u) = \overline{P}_\Omega(u) - \frac{u}{\pi}\int_\Omega\left(\log\frac{1}{\|\vec{y} - \vec{\rho}\|} - \gamma\right)g_\Omega(\vec{\rho}; u)d\vec{\rho}, \tag{5.8}$$

where

$$\overline{P}_\Omega(u) = \lim_{s\downarrow 0} \log \sqrt{2/s}\left(1 - \frac{u}{\pi}\int_\Omega (\log \sqrt{2/s}) G(\vec{\rho};\, u,\, s)\, d\vec{\rho}\right) \tag{5.9}$$

the existence of the limit in (5.9) being part of the assertion.

It follows, in particular, that

$$\frac{u}{\pi}\int_\Omega g_\Omega(\vec{\rho};\, u)\, d\vec{\rho} = 1, \tag{5.10}$$

and

$$g_\Omega(\vec{y};\, u) = P_\Omega(u) - \frac{u}{\pi}\int_\Omega \log \frac{1}{\|\vec{\rho} - \vec{y}\|}\, g_\Omega(\vec{\rho};\, u)\, d\vec{\rho}, \tag{5.11}$$

where

$$P_\Omega(u) = \overline{P}_\Omega(u) + \gamma. \tag{5.12}$$

6. It is a trivial observation that if

$$\Omega' \subset \Omega$$

then

$$g_\Omega(\vec{y};\, u) \leqq g_{\Omega'}(\vec{y};\, u).$$

It is somewhat less trivial but still only an exercise in classical analysis to calculate $g_C(\vec{y};\, u)$ where C is a circle of radius a and center at the origin.

The answer is, for $\vec{y}$ outside the circle,

$$\begin{aligned} g_C(\vec{y};\, u) &= \frac{1 + 2u\int_0^a (\log 1/r) I_0(\sqrt{2u}\, r)\, r\, dr}{2u\int_0^a I_0(\sqrt{2u}\, r)\, r\, dr} \\ &\quad - \frac{1}{2\pi\int_0^a I_0(\sqrt{2u}\, r)\, r\, dr}\int_C\!\!\int \left(\log \frac{1}{\|\vec{\rho} - \vec{y}\|}\right) I_0(\sqrt{2u}\,\|\vec{\rho}\|)\, d\vec{\rho}, \end{aligned} \tag{6.1}$$

where I_0 is again a familiar Bessel function (of first kind and imaginary argument, thus the companion of K_0).

We can use $g_C(\vec{y};\, u)$ as an upper estimate for $g_\Omega(\vec{y};\, u)$ if $C \subset \Omega$. Note now that by comparing (6.1) with (5.11) we have for $\vec{y} \in C$

$$g_C(\vec{y}) = \frac{I_0(\sqrt{2u}\,\|\vec{y}\|)}{2u\int_0^a I_0(\sqrt{2u}\, r)\, r\, dr}. \tag{6.2}$$

Letting $u \uparrow \infty$ we find that

$$\lim_{u \uparrow \infty} g_C(\vec{y}; u) = \log \frac{1}{a} - \log \frac{1}{\|\vec{y}\|}, \qquad \vec{y} \notin C \tag{6.3a}$$

and

$$\lim_{u \uparrow \infty} g_C(\vec{y}; u) = 0, \qquad \vec{y} \in C, \tag{6.3b}$$

and therefore, for the circle at least,

$$U_C(\vec{y}) = \lim_{u \uparrow \infty} g_C(\vec{y}; u) \tag{6.4}$$

is the logarithmic potential of C.

Without going into too much detail, let me merely state that for quite a general class of regions the limit of $g_\Omega(\vec{y}; u)$ as $u \uparrow \infty$ is the logarithmic potential of the region.

If we go back to formula (5.11) we see that it can be rewritten in the form

$$g_\Omega(\vec{y}; u) = P_\Omega(u) - \int_\Omega \log \frac{1}{\|\vec{\rho} - \vec{y}\|} \mu_u(d\vec{\rho}), \tag{6.5}$$

where the measure μ_u is defined by the formula

$$\mu_u(B) = (u/\pi) \int_{B \cap \Omega} g_\Omega(\vec{\rho}; u) d\vec{\rho}, \tag{6.6}$$

and in view of (5.10) we have

$$\mu_u(\Omega) = 1. \tag{6.7}$$

Letting $u \uparrow \infty$ along an appropriate subsequence we have

$$\mu_u(B) \rightarrow \mu(B) \tag{6.8}$$

in the usual sense of convergence of measures and finally

$$U_\Omega(\vec{y}) = \lim_{u \uparrow \infty} g_\Omega(\vec{y}; u) = R(\Omega) - \int \log \frac{1}{\|\vec{\rho} - \vec{y}\|} \mu(d\vec{\rho}), \tag{6.9}$$

where $R(\Omega)$ is clearly the Robin constant of the region Ω.

We also have

$$R(\Omega) = \lim_{u \uparrow \infty} P_\Omega(u) \tag{6.10}$$

although so far the limit is taken along an appropriate subsequence.

7. It is at this point that my story really begins.

Since one knows from classical theory that the equilibrium distribution μ is unique it should be possible to dispense with the proviso that the limit $u \uparrow \infty$ be

taken along an appropriate subsequence.

It should, in particular, be possible to prove (6.10) as it stands.

To do this note first that $P_\Omega(u)$ is bounded for u's bounded away from 0.

To see this integrate (5.11) over Ω obtaining (with the use of (5.10))

$$|\Omega|\,P_\Omega(u) = \frac{\pi}{u} + \frac{u}{\pi}\int_\Omega g_\Omega(\vec{\rho};u)\int_\Omega \log\frac{1}{\|\vec{\rho}-\vec{y}\|}\,d\vec{y}\,d\vec{\rho} \tag{7.1}$$

($|\Omega|$ as usual denotes the Lebesgue measure of Ω). Since clearly there is an M such that

$$\left|\int_\Omega \log\frac{1}{\|\vec{\rho}-\vec{y}\|}\,d\vec{y}\right| < M \tag{7.2}$$

for all $\vec{\rho}\in\Omega$, we have (again using (5.10)) that

$$|\Omega|\,P_\Omega(u) < (\pi/u) + M. \tag{7.3}$$

Next let $\lambda_1, \lambda_2, \lambda_3, \cdots$ be the eigenvalues and $\phi_1, \phi_2, \phi_3, \cdots$ normalized eigenfunctions of the kernel

$$(1/\pi)\log 1/\|\vec{\rho}-\vec{y}\|, \qquad \vec{\rho}, \vec{y}\in\Omega. \tag{7.4}$$

Equations (5.11) and (5.10) yield easily that

$$(P_\Omega(u))^{-1} = \frac{1}{\pi}\sum_{j=1}^{\infty}\frac{\left(\int_\Omega \phi_j(\vec{\rho})\,d\vec{\rho}\right)^2}{(1/u)+\lambda_j} \tag{7.5}$$

and it follows almost at once that

$$\frac{1}{R(\Omega)} = \frac{1}{\pi}\sum_{j=1}^{\infty}\frac{\left(\int_\Omega \phi_j(\vec{\rho})\,d\vec{\rho}\right)^2}{\lambda_j}. \tag{7.6}$$

When I discovered this formula more than fifteen years ago I had already proved a similar formula for the capacity of a three-dimensional region, namely

$$C(\Omega) = \frac{1}{2\pi}\sum_{j=1}^{\infty}\frac{\left(\int_\Omega \phi_j(\vec{\rho})\,d\vec{\rho}\right)^2}{\lambda_j}, \tag{7.7}$$

where the λ's and ϕ's are the eigenvalues and normalized eigenfunctions of the kernel

$$\frac{1}{2\pi}\frac{1}{\|\vec{\rho}-\vec{y}\|}, \qquad \vec{\rho}, \vec{y}\in\Omega. \tag{7.8}$$

The kernel (7.8) is well known to be positive definite but this could clearly not always be true of the kernel (7.4) because the Robin constant can be negative (e.g., for a circle of radius $a>1$). Thus, for some regions at least, not all the λ's could be positive and the question of negative eigenvalues had to be looked into.

Now, one look at (7.5) will show that if $\lambda_j<0$ then either

$$\text{(a)}\ P_\Omega(-1/\lambda_j) = 0 \quad \text{or} \quad \text{(b)}\ \int_\Omega \phi_j(\vec{\rho})d\vec{\rho}) = 0.$$

Now, (b) can be excluded by the following argument: The kernel

$$(1/\pi)K_0(\sqrt{2s}\|\vec{\rho} - \vec{y}\|), \qquad \vec{\rho}, \vec{y} \in \Omega$$

is positive definite for all $s>0$ and therefore we have for *every* $\phi \in L_2(\Omega)$

$$1/\pi \int_\Omega \int_\Omega \phi(\vec{\rho})K_0(\sqrt{2s}\|\vec{\rho} - \vec{y}\|)\phi(\vec{y})d\vec{\rho}d\vec{y} \geqq 0.$$

If however

$$\int_\Omega \phi_j(\vec{\rho})d\vec{\rho} = 0, \tag{7.9}$$

then

$$(1/\pi)\int_\Omega \int_\Omega \phi(\vec{\rho})K_0(\sqrt{2s}\|\vec{\rho} - \vec{y}\|)\phi(\vec{y})d\vec{\rho}d\vec{y}$$

$$= (1/\pi)\int_\Omega \int_\Omega \phi(\vec{\rho})\{K_0(\sqrt{2s}\|\vec{\rho} - \vec{y}\|) - \log\sqrt{2/s} + \gamma\}\phi(\vec{y})d\vec{\rho}d\vec{y} \geqq 0$$

and letting $s \downarrow 0$ we obtain

$$\frac{1}{\pi}\int_\Omega \phi(\vec{\rho}) \log \frac{1}{\|\vec{\rho} - \vec{y}\|} \phi(\vec{y})d\vec{\rho}d\vec{y} \geqq 0. \tag{7.10}$$

In other words, the kernel (7.4) is positive definite on the subspace of $L_2(\Omega)$ defined by (7.9).

If an eigenfunction ϕ_j belonging to a *negative* eigenvalue λ_j did satisfy (7.9) we would have a contradiction since (7.10) would give $\lambda_j \geqq 0$.

Having thus excluded (b) there remains only (a). In other words if $\lambda_j<0$ then $P_\Omega(-1/\lambda_j)=0$.

I shall now prove that $P_\Omega(u)=0$ cannot have two distinct roots. For if u_1 and u_2 were such roots ($u_1 \neq u_2$) then by (5.11) $g_\Omega(\vec{y}; u_1)$ and $g_\Omega(\vec{y}; u_2)$ would be eigenfunctions of (7.4) belonging to distinct eigenvalues and hence *orthogonal* to each other. However, both $g_\Omega(\vec{y}; u_1)$ and $g_\Omega(\vec{y}, u_2)$ are *nonnegative* and hence the only way they could be orthogonal would be if they were to vanish on complementary subsets of Ω. But, if, e.g., $u_2>u_1$, $g_\Omega(\vec{y}; u_2) \leqq g_\Omega(\vec{y}; u_1)$ and $g_\Omega(\vec{y}; u_2)$ would

have to vanish wherever $g(\vec{y}; u_1)$ did, which is clearly impossible.

Thus $P_\Omega(u)=0$ cannot have two distinct roots. It cannot have a multiple root either because by a slight perturbation of the region the degeneracy would be resolved and we would again have a contradiction.

Thus the kernel

$$\frac{1}{\pi}\log\frac{1}{\|\vec{y}-\vec{\rho}\|}, \qquad \vec{y}, \vec{\rho} \in \Omega$$

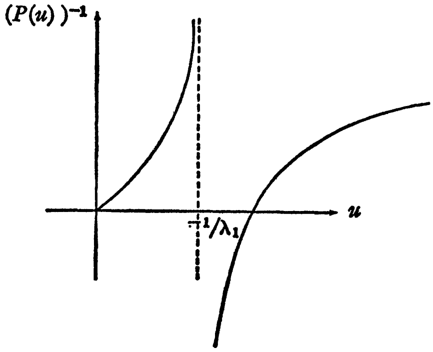
can have at most one negative eigenvalue.

8. It remains to find a convenient criterion for deciding whether there is or there is not a negative eigenvalue.

If $R<0$ there clearly must be a negative eigenvalue as seen by looking at (7.6).

Assume now that $R \geqq 0$ and that e.g., $\lambda_1 < 0$.

It is clear that the graph of $(P(u))^{-1}$ versus u must look like



and that therefore $(P(u))^{-1}$ must vanish for some $u > -1/\lambda_1$ which contradicts the boundedness of $P(u)$.

You could inquire whether it is possible to have an eigenvalue which is equal to 0. The answer is *no* and the proof is easy.

If 0 were an eigenvalue there would exist a nontrivial function $\phi \in L_2(\Omega)$ such that

$$\frac{1}{\pi}\int_\Omega \phi(\vec{y}) \log\frac{1}{\|\vec{\rho}-\vec{y}\|}\,d\vec{y} = 0, \qquad \vec{\rho} \in \Omega. \tag{8.1}$$

By taking the Laplacian of both sides of (8.1) we obtain $\phi=0$ almost everywhere contrary to the assumption.

Summarizing we can say that a necessary and sufficient condition for the kernel (7.4) to have a negative eigenvalue is that the Robin constant be negative.

Since zero is never an eigenvalue we are faced with an amusing situation when we vary the size of the region continuously (keeping, e.g., its shape invariant) so that the Robin constant goes from being negative through zero to being positive. The sole negative eigenvalue does pass through zero (or to be more precise it approaches zero as $R \to 0$) but since zero is not an eigenvalue something strange must happen to the corresponding eigenfunction. In fact it "vanishes" in the sense of mystery stories. However, a careful analysis of Dr. Z. Ciesielski shows that in a certain sense this eigenfunction approaches as $R \to 0$ the equilibrium mass distribution of the region with $R = 0$. So it doesn't really "vanish" but is transfigured.

9. It would be nice if I could tell you that the striking properties of the eigenvalues of the logarithmic kernel (7.4) could be neither found nor proved without the intervention of probability theory.

That indeed they can has been demonstrated quite recently by Dr. J. L. Troutman ([1], [2]) who gave purely classical proofs of the facts concerning the negative eigenvalues of the kernel (7.4). Since I never published my proofs (hoping to include them in a joint book with Dr. Ciesielski but which, for a variety of reasons, will probably never be completed) though I discussed them in a number of lectures, Dr. Troutman's discovery is entirely independent of the whole probabilistic development.

There is, however, a slight "fallout" of the probabilistic approach which perhaps is worthy of mention.

Hidden in the calculations is the heart of the probabilistic proof, namely, that the logarithmic kernel

$$\frac{1}{\pi} \log \frac{1}{\|\vec{r} - \vec{\rho}\|}$$

is a *finite part* of the Laplace transform

$$\int_0^\infty e^{-st} \frac{\exp\left[-\|\vec{r} - \vec{\rho}\|^2/2t\right]}{2\pi t} dt$$

of the basic transition probability density of the plane Brownian motion.

This immediately suggests looking for analogies and the simplest one is the ordinary random walk on a line with equiprobable steps of $+1$ or -1.

Here is a table indicating the analogies:

Plane Brownian Motion	*Simple Random Walk*
Region Ω	Finite set $m_1, m_2, \cdots, m_r$ of integers
$p(\vec{\rho} \mid \vec{r}; t) = \dfrac{\exp(-\|\vec{r} - \vec{\rho}\|^2/2t)}{2\pi t}$	$p(n \mid m; t) = \dfrac{1}{2\pi} \displaystyle\int_0^{2\pi} (\cos^t \xi) \exp[i\xi(m - n)] d\xi$
	$=$ prob. that the displacement in t steps is $m - n$.

Laplace transform

$$\int_0^\infty e^{-st}p(\vec{\rho}\mid\vec{r};t)dt = \frac{1}{\pi}K_0(\sqrt{2s}\|\vec{r}-\vec{\rho}\|)$$

Finite part of the Laplace transform

$$\frac{1}{\pi}\log\frac{1}{\|\vec{r}-\vec{\rho}\|}$$

Logarithmic kernel over Ω.

Generating function

$$\sum_0^\infty z^t p(n\mid m;t) = \frac{1}{2\pi}\int_0^{2\pi}\frac{\cos(m-n)\xi}{1-z\cos\xi}d\xi$$

Finite part of the generating function

$$-\frac{1}{2\pi}\int_0^{2\pi}\frac{1-\cos(m-n)\xi}{1-\cos\xi}d\xi = -\left|m-n\right|$$

Matrix $-((|m_i-m_j|)),\quad 1\leq i,j\leq r.$

It is now obvious what the theorem is.

The matrix $-((|m_i-m_j|))$ has at most one negative eigenvalue and since it is finite and has trace zero it must have *exactly* one negative eigenvalue.

Integers m_i can be replaced by rational numbers and then, by continuity, by real numbers.

We have thus finally the following theorem (see [3]):

If $a_1, a_2, \cdots, a_r$ are distinct real numbers then the matrix $((|a_i-a_j|))$ has exactly one positive eigenvalue.

This theorem can also be proved in a more direct way (as was shown to me some years ago by Professor Pólya) but you must agree with me that only through probability theory does one see that this theorem and the theorem that the logarithmic kernel has at most one negative eigenvalue are really one and the same. And this, I hope you will again agree with me, is quite remarkable.

References

1. J. L. Troutman, The logarithmic potential operator, Ill. J. of Math., 11 (1967) 360–374.
2. ———, The logarithmic eigenvalues of plane sets, *ibid.* 13 (1968) 95–107.
3. M. Kac, A class of limit theorems, Trans. Amer. Math. Soc., 84 (1957) 459–471, esp. pp. 470–471.

GEOMETRIC EMBEDDING OF COMPLEXES

R. A. DUKE, University of Washington

Introduction. It is well known that each finite n-dimensional simplicial complex can be embedded in the $(2n+1)$-dimensional Euclidean space E^{2n+1} and is, in fact, geometrically embeddable there. (A simplicial complex is *geometrically embeddable* in E^d provided that there exists an embedding such that each of its simplices is a convex, rectilinear simplex in E^d.) This result is the best possible in the sense that for each integer $n\geq 1$ there exists a finite n-dimensional simplicial complex which is not even *topologically embeddable* in E^{2n}, that is, not homeomorphic to any subset of E^{2n}. It is also known that for $k<2n$ an n-dimensional simplicial complex which is topologically embeddable in E^k is not necessarily geometrically embeddable there.

Beyond this, however, little is known concerning either the topological or

geometric embeddability of specific finite-dimensional simplicial complexes in finite-dimensional Euclidean spaces. In Section 2 we will review some of the established results as well as state some conjectures which have been made concerning the embeddability of such complexes.

Many questions remain unanswered even for the case of 2-dimensional complexes. One such deals with triangulations of 2-manifolds. By a *triangulation* of a manifold M we shall mean a finite simplicial complex homeomorphic to M. Each triangulation of an orientable 2-dimensional manifold is topologically embeddable in E^3, and a theorem of Steinitz [**18**] concerning convex 3-polytopes has as a corollary that each triangulation of the 2-sphere can even be embedded geometrically in E^3. Grünbaum [**10**, Chpt. 13] has asked whether geometric embeddings in E^3 exist for every triangulation of each orientable 2-manifold. In Section 3 we will discuss the results which have been obtained concerning this more restricted, yet apparently quite difficult problem, which is still open even for the case of the torus.

2. General Results. Henceforth *n-complex* will mean a finite, n-dimensional, simplicial complex. The geometric embeddability of an arbitrary n-complex C in E^{2n+1} follows from the fact that the images of the vertices of C may be chosen to be a set of points which are in general position in E^{2n+1}, that is, a set of points in E^{2n+1} such that each subset of at most $2n+2$ members is affinely independent. For in this case, the image of each simplex of C may be taken to be the convex hull of the images of those vertices which it contains. Since C is of dimension n, none of its simplices has more than $n+1$ vertices, and for any pair of simplices of C there exists a larger simplex in E^{2n+1} (not necessarily in the image of C) having the images of the given pair among its faces. Thus the images of any two simplices of C are either disjoint or intersect in a simplex which is a common face of both. (This result has been generalized by M. A. Perles to include certain complexes whose cells are not necessarily simplices [see **10**, p. 204].)

The first examples of n-complexes not embeddable in E^{2n} are due to van Kampen [**13**] and Flores [**8, 9**]. In order to describe these examples as well as generalizations due to Grünbaum [**10, 11**] and Wu [**20**], we will need the following additional terms. The *complete n-complex with k-vertices*, $C(n, k)$, is the n-dimensional simplicial complex having k vertices and with the property that each subset of at most $n+1$ of its vertices determines a simplex in the complex. The *join* of two disjoint simplices T and S having vertices $v_1, \cdots, v_n$ and $w_1, \cdots, w_m$ respectively is the simplex having $\{v_1, \cdots, v_n, w_1, \cdots, w_m\}$ as its set of vertices, and is written $T \vee S$. For disjoint simplicial complexes K and C, the join $K \vee C$ is the complex consisting of all simplices of the type $T \vee S$, where T is a simplex of K and S a simplex in C. Here the empty set $\emptyset$ is considered to be a (-1-dimensional) simplex in each complex. If K is an n-complex and C an m-complex, then $K \vee C$ is an $(n+m+1)$-complex. The operation of forming the join of disjoint complexes is both associative and commutative in the sense that for any pairwise disjoint complexes K_1, K_2, and K_3, there exist isomorphisms between

$K_1 \vee K_2$ and $K_2 \vee K_1$ and between $(K_1 \vee K_2) \vee K_3$ and $K_1 \vee (K_2 \vee K_3)$. Using this terminology we have the following result.

THEOREM 1. *Let n_i, m_i, and p be nonnegative integers such that $n_i+3 \leqq m_i \leqq 2n_i+3$, for $i=1, 2, \cdots, p$. Then the $(n_1+n_2+\cdots+n_p+p-1)$-complex*

$$C(n_1, m_1) \vee C(n_2, m_2) \vee \cdots \vee C(n_p, m_p)$$

is not topologically embeddable in Euclidean $(m_1+m_2+\cdots+m_p-p-2)$-space, but is even geometrically embeddable in Euclidean $(m_1+\cdots+m_p-p-1)$-space.

The proof of this theorem, patterned on the original treatment of Flores, is rather long and makes use of his notion of "self-entangled complexes" and the Borsuk-Ulam theorem on antipodes (see [11]).

In particular, if $n=n_1+\cdots+n_p+p-1$ and $m_i=2n_i+3$ for $i=1, 2, \cdots, p$, the n-complexes obtained are not embeddable in E^{2n}. (For $n=1$, one obtains the two "Kuratowski skew curves" which can be used to characterize nonplanar graphs. For $p=1$ and $m_1=2n_1+3$, we have the result of Flores and van Kampen that the complex $C(n, 2n+3)$, which may be interpreted as the n-skeleton of the $(2n+2)$-dimensional simplex, is not topologically embeddable in E^{2n}.)

Concerning the n-complexes with $m_i=2n_i+3$ for $i=1, 2, \cdots, p$, Grünbaum [11] has proved the following (see also [21]):

THEOREM 2. *Let n, p, $n_1, \cdots, n_p$ be nonnegative integers such that $n=n_1+\cdots+n_p+p-1$. Then every proper subcomplex of the simplicial n-complex*

$$C(n_1, 2n_1+3) \vee C(n_2, 2n_2+3) \vee \cdots \vee C(n_p, 2n_p+3)$$

is geometrically embeddable in E^{2n}.

The "minimality" property of this theorem does not hold for all of the complexes of Theorem 1 for which $m_i < 2n_i+3$ for some i. These results suggest the following, also due to Grünbaum.

CONJECTURE. *Each finite n-dimensional simplicial complex which is topologically embeddable in E^{2n} is also geometrically embeddable there.*

For $k < 2n$, an n-complex which is topologically embeddable in E^k need not be geometrically embeddable there. In [4], Cairns gave an example of a 3-complex homeomorphic to the 3-simplex but not geometrically embeddable in E^3. This example was modified by van Kampen in [14]. In connection with the enumeration of a certain class of convex 4-polytopes, Grünbaum and Sreedharan [12] discovered a 3-complex M which is topologically equivalent to the 3-sphere and which has the property that its 2-skeleton C, that is, the complex C consisting of all 0, 1, and 2-dimensional simplices of M, is not geometrically embeddable in E^3. Barnette [2] has found a second 3-sphere with similar properties. By taking joins of the form $C \vee \{v_1\} \vee \cdots \vee \{v_{k-2}\}$, where C is the 2-skeleton of M and each v_i is a single point, one can establish the following:

THEOREM 3. *For each $k \geqq 2$, there exists a k-complex which is topologically embeddable in E^{k+1}, but which is not geometrically embeddable in E^{k+1}.*

That no such complexes exist for $k=1$ is a result of the well-known theorem on the straight line representation of planar graphs due to Wagner [19] and Fáry [7].

In the most general form, the question raised by these results might be put as follows. For what integers k, n, and m does there exist a simplicial k-complex which is topologically embeddable in E^n but which is not geometrically embeddable in E^m. Solutions for the cases $k+2=n=m$ and $m=n+1$ would be of particular interest, even for $k=2$.

By taking the "dual" complex (see [10]) of the 3-sphere M, one obtains a 3-dimensional cell-complex M^*, not a simplicial complex, which is also homeomorphic to the 3-sphere and is such that neither M^* nor its 2-skeleton is geometrically embeddable in any finite-dimensional Euclidean space. (M^* was first described by Brückner [3] who listed it, erroneously, among the boundary complexes of certain convex 4-dimensional polytopes.) The existence of complexes such as M^* makes the following conjecture even more interesting.

CONJECTURE. *For each integer $n \geqq 1$, every finite, simplicial n-dimensional complex homeomorphic to the n-dimensional sphere is geometrically embeddable in E^{n+1}.*

3. Triangulated 2-manifolds. As noted above, little is known concerning even the geometric embeddability of 2-complexes. Some information can be derived in this case, however, from a theorem of Steinitz [18] on convex 3-polytopes. A subset P of a finite-dimensional Euclidean space is called a *convex polytope*, provided that P is the intersection of a finite number of closed halfspaces and is bounded. In this case P is the convex hull of a finite set of points, its *vertices*. The *graph of P* is the 1-dimensional complex consisting of the vertices and edges of P. The theorem of Steinitz, one of the most important results in the study of convex polytopes, implies that an abstract 1-complex G is the graph of some 3-dimensional convex polytope if and only if G is planar and 3-connected (that is, G can be embedded in E^2 and has at least 4 vertices, any two of which can be joined by at least 3 independent paths). It follows from this theorem that if C is any 2-dimensional cell-complex whose point set is homeomorphic to the 2-sphere, then there exists a one-to-one incidence-preserving correspondence between C and the boundary complex consisting of the proper faces of some 3-dimensional convex polytope. Hence such a complex C is geometrically embeddable in E^3. (Compare this result with the case of cell-complexes homeomorphic to the 3-sphere, which includes the complex M^* of Section 2.)

The question then arises whether each 2-dimensional cell-complex homeomorphic to some closed, connected, orientable 2-manifold is geometrically embeddable in E^3. In his book on convex polytopes [10, Exercise 13.2.3], Grünbaum has pointed out that the answer to this question is negative since a 2-dimensional cell-complex homeomorphic to such a manifold M and such that each

vertex is the endpoint of exactly 3 edges is geometrically embeddable in E^3 if and only if M is the 2-sphere. Each *simplicial* 2-complex which is a triangulation of an orientable 2-manifold is topologically embeddable in E^3 (see [**18**]), and Grünbaum has asked whether geometric embeddings in E^3 exist for all such *triangulations*. Even when restricted to the case of the torus, the orientable 2-manifold of lowest positive genus, this question seems to be quite difficult.

An easy argument involving the Euler characteristic shows that a triangulation of the torus must have at least 7 vertices and it can be checked that there exists only one such triangulation, T_7, having exactly this many vertices. In T_7 each two distinct vertices are joined by an edge. It has been known for some time that a geometric embedding of the 2-complex T_7 in E^3 is possible. Möbius [**16**] described this complex as an example of a "polyhedron of the second class" (that is, of genus 1) and stated without proof that it could be constructed by fitting together 7 tetrahedra. A complete description of a geometric embedding has been given by Császár [**6**].

Another geometric embedding of T_7 can be obtained by making use of the notion of a *cyclic polytope*. A cyclic d-dimensional polytope is a convex polytope with $d+1$ or more vertices each lying on the *moment curve* M_d in E^d, where $M_d=\{(t, t^2, \cdots, t^d):t \text{ a real number}\}$. The cyclic $2d$-polytopes provide the simplest examples of *d-neighborly polytopes*, where a polytope P is said to be n-neighborly provided that each subset of n vertices of P is the set of vertices of an $(n-1)$-dimensional face of P (which must therefore be an $(n-1)$-dimensional simplex). (The existence of neighborly polytopes, although rediscovered only recently, appears to have been known to Carathéodory [**5**] in 1911 and, in E^4, to Brückner [**3**] in 1909. For more on their interesting history see Klee [**15**].) The cyclic 4-polytope P_7 with 7 vertices is such that its 2-skeleton contains a subcomplex isomorphic to T_7. Thus by selecting any 3-dimensional face F of a copy of P_7 in E^4 and projecting P_7 onto the affine hull of F from a point of E^4 not in P_7 but sufficiently close to the relative interior of F, one obtains a 3-complex, a *Schlegel diagram* of P, which is geometrically embedded in E^3 and combinatorially equivalent to the boundary complex of P_7 with F deleted. This Schlegel diagram contains a copy of T_7 geometrically embedded in E^3.

In the same way, any triangulation of a 2-manifold which is contained in the 2-skeleton of a 4-polytope in this sense could be embedded geometrically in E^3. While for each nonnegative integer g there exists a convex 4-polytope which contains a triangulation of a 2-manifold of genus g, it can be shown [**1**] that the only triangulated manifolds contained in the 2-skeleton of a *cyclic* 4-polytope are the 2-sphere and the torus. Moreover, a study of the structure of the cyclic polytopes shows that a triangulation of the torus is contained in a cyclic 4-polytope if and only if each has the same number of vertices and there exists a simple closed path through the vertices of the triangulation such that every triangle of the triangulation has exactly one edge on this path. This criterion can be used to show that one of the 7 triangulations of the torus with 8 vertices can not be found in a cyclic 4-polytope. It is not known, however, whether each triangula-

tion of the torus or even each orientable 2-manifold can be found in 4-polytope, perhaps always in some neighborly 4-polytope. Altshuler [1] has used the simple closed path criterion to show that each triangulation of the torus in which each vertex is an endpoint of exactly 6 edges can be embedded geometrically E^3, and has considered another special class of 4-polytopes, the "stacked polytopes."

Another approach, suggested by a proof of a special case of the theorem of Steinitz (see [10, Exercise 13.1.3]), consists of reducing the list of all triangulations of the torus to certain "minimal" triangulations by successively eliminating certain pairs of triangular 2-cells (and hence also some vertices and edges). By such a procedure one can obtain a finite list of minimal triangulations of the torus, each with 10 or fewer vertices. It can be shown, both directly and from the fact that each is contained in some convex 4-polytope (although not necessarily a cyclic one), that these minimal triangulations can be embedded geometrically in E^3. Attempts to use these facts to obtain the result for all triangulations of the torus, however, have been unsuccessful.

Barnette (unpublished) has shown that geometric embeddings exist for a large number of triangulations which are obtainable by subdividing certain non-simplicial 2-complexes homeomorphic to the torus. He has also established the existence of geometric embeddings in E^3 for all triangulations of the "pinched torus," that is, those 2-complexes resulting from the identification of two non-adjacent vertices of a triangulated 2-sphere.

It appears that no counterexamples are known to the following conjecture which includes the last conjecture of Section 2 as well as the principal question of this section.

CONJECTURE. *For any positive integers k and n, each triangulated k-dimensional manifold which is topologically embeddable in E^n is geometrically embeddable there.*

References

1. A. Altshuler, Polyhedral realization in R^3 of triangulations of the torus and 2-manifolds in convex polytopes, Ph.D. thesis, The Hebrew University, Jerusalem, 1969.

2. D. Barnette, Diagrams and Schlegel diagrams, Proc. Intern. Conf. on Combinatorial Structures and their Applications, June, 1969, Calgary (to appear).

3. M. Brückner, Über die Ableitung der allgemeinen Polytope und die nach Isomorphismus verschiedenen Typen der allgemeinen Achtzelle (Oktatope), Verh. Kon. Akad. Wet. Amsterdam, (Eerste Sect.), Deel X. No. 1 (1909) 29 pages+2 plates.

4. S. S. Cairns, Triangulated manifolds which are not Brouwer manifolds, Ann. Math., 41 (1940) 792–795.

5. C. Carathéodory, Über den Variabilitätsbereich der Fourierschen Konstanten von positiven harmonischen Funktionen, Rend. Circ. Mat., Palermo, 32 (1911) 193–217.

6. A. Császár, A polyhedron without diagonals, Acta. Sci. Math. (Szeged), 13 (1949) 140–142.

7. I. Fáry, On straight line representation of planar graphs, Acta. Sci. Math. (Szeged), 11 (1948) 229–233.

8. A. Flores, Über die Existence n-dimensionaler Komplexe, die nicht in den R_{2n} topologisch einbettbar sind, Ergeb. Math. Kolloq., 5 (1932/33) 17–24.

9. ———, Über n-dimensionaler Komplexe, die im R_{2n+1} absolut selbstverschlungen sind, Ergeb. Math. Kolloq., 6 (1933/34) 4–6.

10. B. Grünbaum, Convex Polytopes, Wiley, New York, 1967.

11. ———, Imbeddings of Simplicial Complexes, Comment. Math. Helv. (to appear).

12. B. Grünbaum and V. P. Sreedharan, An enumeration of simplicial 4-polytopes with 8 vertices, J. Combinatorial Theory, 2 (1967) 437–465.

13. E. R. van Kampen, Komplexe in euklidischen Räumen, Abh. Math. Sem. Univ. Hamburg, 9 (1932) 72–78, 152–153.

14. ———, Remark on the address of S. S. Cairns, "Lectures in topology," University of Michigan Conference 1940, edited by R. L. Wilder and W. L. Ayres, pp. 311–313, Ann Arbor, 1941.

15. V. Klee, Convex polytopes and linear programming, Proc. IBM Sci. Comp. Symp. on Combinatorial Problems, March 1964, pp. 123–158 (1966). (Also Mathematical Note No. 366, Boeing Scientific Research Laboratories, Seattle, Washington, 1964.)

16. A. F. Möbius, Über die Bestimmung des Inhalts eines Polyëders, Berichten der K. Sächs. Ges. Wiss., Math.-Phys. Klasse, 17 (1865) 31–68. (Also in Gesammelte Werke, Vol. 2, Hirzel, Leipzig, 1886.)

17. H. Seifert and W. Threlfall, Lehrbuch der Topologie, Chelsea, New York, 1947.

18. E. Steinitz and H. Rademacher, Vorlesungen über die Theorie der Polyeder, Springer, Berlin, 1934.

19. K. Wagner, Bemerkungen zum Vierfarbenproblem, Jber. Deutsch. Math.-Ver., 46 (1936) 26–32.

20. W.-T. Wu, A Theory of Imbedding, Immersion, and Isotopy of Polytopes in a Euclidean Space, Science Press, Peking, 1965.

21. J. Zaks, On a minimality property of complexes, Proc. Amer. Math. Soc., (to appear).

AN ANALYTIC PROOF OF YOUNG'S INEQUALITY

J. B. DIAZ, Rensselaer Polytechnic Institute AND
F. T. METCALF, University of California, Riverside

The main result of the paper is the analytic proof of Young's inequality, given in section 2. While the proof in section 2 may be read entirely independently of the considerations of section 1, the discussion in section 1 is essential for a complete overall understanding of the subject, both historically and mathematically.

1. Introduction (or, Geometry versus Analysis).

1.A. The inequality referred to in the title of this paper appears, for the first time, on page 226 of W. H. Young [1], who expresses himself as follows: "If

$$v = U(u)$$

is a positive monotone increasing function of a positive real variable u, possessing a differential coefficient $U'(u)$ which is positive (>0), and $V(v)$ is the inverse function such that

$$u = V(v),$$

then

$$uv - ab \leqq \int_a^u U(z)dz + \int_b^v V(z)dz, \tag{1}$$

where

$$b = U(a).$$

"In fact the function of u represented by $uv - ab - \int_a^u U(z)dz$ has a negative second differential coefficient, while its first differential coefficient vanishes when $v = U(u)$, that is when $u = V(v)$. It has accordingly a maximum whose value is

$$vV(v) - ab - \int_{V(b)}^{V(v)} U(z)dz, \qquad \text{that is} \int_{V(b)}^{V(v)} zU'(z)dz;$$

or, what is the same thing, $\int_b^v V(z)dz$. Hence, the inequality in question follows."

Two comments, relative to this quotation of W. H. Young, are appropriate at this point:

In the first place, it is clear that, in order for Young's argument to hold, one must suppose something else about the derivative $U'(z)$, besides merely that it exists and is positive. Because, if nothing further is assumed, then the definite integral $\int_{V(b)}^{V(v)} zU'(z)dz$, which is used in Young's proof, need not exist at all.

In the second place, if the derivative $U'(z)$ is assumed, *further*, to be continuous (just existence is *not* enough), then one can write, with integration in the Riemann-Stieltjes sense, that

$$\int_{V(b)}^{V(v)} zU'(z)dz = \int_{V(b)}^{V(v)} zdU.$$

In view of this, it seems clear how later writers (see below) were led to consider Stieltjes integration in their proofs of Young's inequality.

1.B. The inequality of W. H. Young [**1**] is given in the book of Hardy, Littlewood, and Pólya [**2**, p. 111] as follows:

"Suppose that $\phi(0) = 0$, and that $\phi(x)$ is continuous and strictly increasing for $x \geqq 0$; that $\psi(x)$ is the inverse function, so that $\psi(x)$ satisfies the same conditions; and that $a \geqq 0$, $b \geqq 0$. Then

$$ab \leqq \int_0^a \phi(x)dx + \int_0^b \psi(x)dx.$$

There is equality only if $b = \phi(a)$.

"The theorem becomes intuitive if we draw the curve $y = \phi(x)$ or $x = \psi(y)$, and the lines $x = 0$, $x = a$, $y = 0$, $y = b$, and consider the various areas bounded by them. A formal proof is included in that of the more general theorems which follow."

The formal proof alluded to, in this quotation from Hardy, Littlewood, and

Pólya [2], which is contained on pp. 112–113 of [2], involves several, at first glance, seemingly extraneous notions; for example, the notion of a function of bounded variation, and of Riemann-Stieltjes integration (in particular, the formula for integration by parts in Riemann-Stieltjes integrals). The same remark applies to the proof given in E. C. Francis and J. E. Littlewood [3, pp. 47–48]. (Added October 31, 1969: The same remark also applies to the proof given on p. 70 of the recent book *Lectures on Measure and Integration*, by H. Widom, to which our attention was kindly drawn by the referee.)

1.C. In his book on trigonometrical series, Zygmund [4, pp. 64–65] dismisses any analytic attempt at proof, as follows: "Let $\phi(u)$, $u \geqq 0$, $\psi(v)$, $v \geqq 0$, be two functions, continuous, vanishing at the origin, strictly increasing, tending to ∞, and inverse to each other. Then, for $a, b \geqq 0$, we have the inequality, due to Young,

$$(1) \qquad ab \leqq \Phi(a) + \Psi(b), \qquad \text{where } \Phi(x) = \int_0^x \phi du, \qquad \Psi(y) = \int_0^y \psi dv.$$

The geometrical proof is obvious. It is also easy to see that the sign $\leqq$ can be replaced by $=$ if and only if $b = \phi(a)$."

The easily constructed geometrical pictures, which form the basis for such a proof, can be found, for example, in Tolsted [5].

1.D. In his book on integration, McShane [6, pp. 131–132, p. 216], seemingly unwilling to rely on a purely geometric proof, gives an analytic proof of Young's inequality, patterned after the original proof of Young, quoted above; which, however, in its best version, involves the extra assumption that ϕ is absolutely continuous on every finite interval. (For another comment on McShane's analytic proof, see the remark at the end of section 2.)

1.E. Geometric proofs of mathematical propositions may sometimes lead one astray; compare, for example, the well-known geometric "proof" that every plane triangle is isosceles. Therefore, no matter how convincing any geometric proof of Young's inequality may be, it is, nevertheless, essential to have a valid analytic proof. However, the construction of a valid analytic proof (see some of the references quoted above) does not appear, by far, to be as trivial a matter as the construction of a geometric proof. Be that as it may, it is the purpose of the present paper to provide a direct analytic proof of Young's inequality, valid for *continuous* functions ϕ, which is not beset with the difficulties of Riemann-Stieltjes integration, absolute continuity on every finite interval, etc., which were mentioned in the preceding discussion. Moreover, the proof given here lends itself to generalizations of Young's inequality, as will be developed in full elsewhere.

2. Analytic Proof of Young's Inequality.

The aforementioned analytic derivation of Young's inequality will now be presented (f now replaces the ϕ of section 1).

THEOREM. *Suppose that f is a real-valued function defined for all* $0 \leqq t < \infty$, *and is such that*

(i) $f(0)=0$,

(ii) *f is continuous for* $0 \leqq t < \infty$ *(from the right, at* $t=0$*)*,

(iii) *f is strictly increasing for* $0 \leqq t < \infty$, *and* $\lim_{t\to+\infty} f(t) = +\infty$.

(Under these circumstances, f has an inverse function f^{-1}, which is defined for $0 \leqq t < \infty$, *and which obeys* (i), (ii,) *and* (iii), *with f replaced by f^{-1}, of course.)*

Then, for any $a \geqq 0$, $b \geqq 0$, *one has*

$$ab \leqq \int_0^a f(t)dt + \int_0^b f^{-1}(t)dt,$$

with equality if and only if $b=f(a)$.

Proof. Define, for simplicity in writing, for $0 \leqq a < \infty$ and $0 \leqq b < \infty$,

$$\Delta(a, b) = \int_0^a f(t)dt + \int_0^b f^{-1}(t)dt - ab.$$

Then, the inequality to be shown is just $0 \leqq \Delta(a, b)$.

(I). It will first be shown that

$$\Delta(a, f(a)) \leqq \Delta(a, b)$$

for $0 \leqq a < \infty$ and $0 \leqq b < \infty$, with equality if and only if $b=f(a)$. For any such a and b, one has

$$\Delta(a, b) - \Delta(a, f(a)) = \int_{f(a)}^b [f^{-1}(t) - a]dt = \int_b^{f(a)} [a - f^{-1}(t)]dt.$$

There are two cases to consider. The first case is $b \geqq f(a)$. Here, whenever $b \geqq t \geqq f(a)$, one has $f^{-1}(b) \geqq f^{-1}(t) \geqq f^{-1}[f(a)] = a$. Consequently,

$$\Delta(a, b) - \Delta(a, f(a)) = \int_{f(a)}^b [f^{-1}(t) - a]dt \geqq 0.$$

Since $f^{-1}(t) - a$ is continuous, and strictly increasing, for $f(a) \leqq t \leqq b$, equality will hold if and only if $b=f(a)$. The second case is $b \leqq f(a)$. Here, whenever $b \leqq t \leqq f(a)$, one has $f^{-1}(b) \leqq f^{-1}(t) \leqq f^{-1}[f(a)] = a$. Consequently,

$$\Delta(a, b) - \Delta(a, f(a)) = \int_b^{f(a)} [a - f^{-1}(t)]dt \geqq 0.$$

Since $a - f^{-1}(t)$ is continuous, and strictly decreasing, for $b \leqq t \leqq f(a)$, equality will hold if and only if $b=f(a)$.

(II). Putting, for brevity,

$$\delta(a) = \Delta(a, f(a)) = \int_0^a f(t)dt + \int_0^{f(a)} f^{-1}(t)dt - af(a),$$

for $0 \leqq a < \infty$, it will now be shown that

$$\delta(a) = 0$$

for all $0 \leqq a < \infty$ (this will be done by showing that the function δ, which is continuous for $0 \leqq a < \infty$, possesses a derivative which is zero for all $0 < a < \infty$; while, obviously, $\delta(0) = 0$). Then, for any $a > 0$, and $-a < \epsilon < \infty$, one has

$$\begin{aligned}\delta(a + \epsilon) - \delta(a) &= \int_a^{a+\epsilon} [f(t) - f(a + \epsilon)]dt + \int_{f(a)}^{f(a+\epsilon)} [f^{-1}(t) - a]dt \\ &= \int_{a+\epsilon}^{a} [f(a + \epsilon) - f(t)]dt + \int_{f(a+\epsilon)}^{f(a)} [a - f^{-1}(t)]dt.\end{aligned}$$

The cases $0 < \epsilon < \infty$ and $-a < \epsilon < 0$ must be considered separately:

Case (i): $0 < \epsilon < \infty$. In this case, it follows that, since the functions f and f^{-1} are strictly increasing,

$$\begin{aligned}\delta(a + \epsilon) - \delta(a) \geqq &\int_a^{a+\epsilon} [f(a) - f(a + \epsilon)]dt \\ &+ \int_{f(a)}^{f(a+\epsilon)} [f^{-1}(f(a)) - a]dt = -\epsilon[f(a + \epsilon) - f(a)];\end{aligned}$$

and also that

$$\begin{aligned}\delta(a + \epsilon) - \delta(a) \leqq &\int_a^{a+\epsilon} [f(a + \epsilon) - f(a + \epsilon)]dt \\ &+ \int_{f(a)}^{f(a+\epsilon)} [f^{-1}(f(a + \epsilon)) - a]dt = \epsilon[f(a + \epsilon) - f(a)].\end{aligned}$$

Case (ii): $-a < \epsilon < 0$. In this case, it follows that

$$\begin{aligned}\delta(a + \epsilon) - \delta(a) \geqq &\int_{a+\epsilon}^{a} [f(a + \epsilon) - f(a)]dt \\ &+ \int_{f(a+\epsilon)}^{f(a)} [a - f^{-1}(f(a))]dt = -\epsilon[f(a + \epsilon) - f(a)];\end{aligned}$$

and also that

$$\begin{aligned}\delta(a + \epsilon) - \delta(a) \leqq &\int_{a+\epsilon}^{a} [f(a + \epsilon) - f(a + \epsilon)]dt \\ &+ \int_{f(a+\epsilon)}^{f(a)} [a - f^{-1}(f(a + \epsilon))]dt = \epsilon[f(a + \epsilon) - f(a)].\end{aligned}$$

Thus, for any $a > 0$, and $-a < \epsilon < \infty$, with $\epsilon \neq 0$, one obtains

$$-\left| f(a+\epsilon) - f(a) \right| \leqq \frac{\delta(a+\epsilon)-\delta(a)}{\epsilon} \leqq \left| f(a+\epsilon) - f(a) \right| .$$

Since f is continuous at a, it follows that

$$\lim_{\epsilon\to 0} \frac{\delta(a+\epsilon)-\delta(a)}{\epsilon} = 0$$

for $0<a<\infty$. In other words, $\delta'(a)=0$ for $0<a<\infty$; while, obviously, $\delta(0)=0$. This means that $\delta(a)=0$ for all $0\leqq a<\infty$.

Thus, one has completed the proof that

$$\Delta(a, b) \geqq \Delta(a, f(a)) = 0,$$

with equality if and only if $b=f(a)$, as desired.

REMARK 1: Part (I) of the preceding proof, that

$$\Delta(a, f(a)) \leqq \Delta(a, b),$$

is immediate, for any continuous function f. However, part (II) of the preceding proof, which involves showing that the derivative of the function

$$\delta(a) = \Delta(a, f(a)) = \int_0^a f(t)dt + \int_0^{f(a)} f^{-1}(t)dt - af(a)$$

exists and is identically zero for $0<a<\infty$, is not quite so obvious, for a continuous function f. This is because, for a function f which is merely continuous (rather than continuously differentiable), it is not possible to differentiate directly, term by term, using the so-called "Leibniz rule for differentiating definite integrals with variable limits of integration," in order to obtain the derivative $\delta'(a)$. It is for this reason, of being able to differentiate term by term in obtaining $\delta'(a)$, that McShane [6, pp. 131–132] assumes first that f is continuously differentiable. Later in his book, McShane [6, p. 216; see also footnote * on p. 132] indicates that his proof can be extended to functions f which are absolutely continuous on every finite closed interval. But, since there are strictly increasing continuous functions which are not absolutely continuous, McShane's proof does not cover the case of an arbitrary continuous f. (For an example of such a function, simply "add x" to the "Cantor function" discussed in McShane [6, pp. 48–49].)

It is also true that one could start with the argument of McShane [6, pp. 131–132] for continuously differentiable f, and seek to extend it to cover the case of functions f which are merely continuous, by employing an approximation theorem (for example, Weierstrass' approximation theorem of continuous functions by polynomials), which guarantees that any continuous f is the limit of a sequence of continuously differentiable functions. But, all in all, one has the feeling that, after everything is thus said and done, the resulting line of reason-

ing will certainly not possess the strictly elementary character of the proof given in this paper.

REMARK 2: It is not strictly necessary to require that $\lim_{t\to+\infty} f(t)=+\infty$, as is done in hypothesis (iii) of the theorem; this is required here only to simplify somewhat the writing in the proof. In case $\lim_{t\to+\infty} f(t)$ is not infinite, but positive, then the inverse function f^{-1} is defined only on the finite interval $0\leqq b<\lim_{t\to+\infty} f(t)$, and the same proof as given above applies, except that, both in the statement of the theorem and in the proof, the inequality $0\leqq b<+\infty$ must be replaced, wherever it appears, by the clumsier looking inequality $0\leqq b<\lim_{t\to+\infty} f(t)$. Thus, the proof given above, suitably interpreted, covers fully both cases, that of $\lim_{t\to+\infty} f(t)=+\infty$, and that of $\lim_{t\to+\infty} f(t)$ positive.

REMARK 3: The letters Δ and δ, employed in the course of the proof, are, really, when one comes right down to it, entirely superfluous; these letters are merely introduced in order to simplify the writing of the argument. In order to make evident the simplification of the writing involved, it is, perhaps, useful to state explicitly that, in part (I), one is seeking to show that

$$\int_0^a f(t)dt+\int_0^{f(a)} f^{-1}(t)dt - af(a) \leqq \int_0^a f(t)dt+\int_0 f^{-1}(t)dt - ab,$$

for all $a\geqq 0$, $b\geqq 0$, with equality if and only if $b=f(a)$; whereas, in part (II), one is seeking to show that

$$0=\int_0^a f(t)dt+\int_0^{f(a)} f^{-1}(t)dt - af(a),$$

for all $a\geqq 0$.

Furthermore, it is to be noticed that these two parts of the analytic proof correspond exactly to the two types of geometric pictures which are drawn in the geometric "proof" of Young's inequality: Part (I) corresponds to the figure drawn when $b\neq f(a)$, while part (II) corresponds to the figure drawn when $b=f(a)$.

References

1. W. H. Young, On classes of summable functions and their Fourier series, Proc. Royal Soc., Series (A), 87 (1912) 225–229.

2. G. H. Hardy, J. E. Littlewood, and G. Pólya, Inequalities, Cambridge, 1934.

3. E. C. Francis and J. E. Littlewood, Examples in infinite series, Cambridge, 1928.

4. A. Zygmund, Trigonometrical series, 2nd edition, Chelsea, New York, 1952.

5. Elmer Tolsted, An elementary derivation of the Cauchy, Hölder, and Minkowski inequalities from Young's inequality, Math. Mag., 37 (1964) 2–12.

6. E. J. McShane, Integration, Princeton University Press, 1947.

EDITED BY DAVID DRASIN

Manuscripts for this Department should be sent to David Drasin, Division of Mathematical Sciences. Purdue University, Lafayette, IN 47907.

ON A RESULT OF LIBRI AND LEBESGUE

K. S. WILLIAMS, Carleton University, Ottawa

1. Introduction. In this note we observe how a simple property of a primitive nth root of unity provides us with a counting function for the number of solutions of a congruence $f(x_1, \cdots, x_k) \equiv 0 \pmod{n}$. We illustrate the idea by taking $f(x_1, \cdots, x_k) = x_1^l + \cdots + x_k^l$ and a prime $n = p \equiv 1 \pmod{l}$. We are led naturally to the q-nomial periods of the pth roots of unity where $q = (p-1)/l$ [2]. We express the number of solutions of $x_1^l + \cdots + x_k^l \equiv 0 \pmod{p}$ in terms of these periods, rediscovering an old result due to Libri and Lebesgue [1]. An alternative form of this result is also proved which provides a generalization of one due to the author when $l = 3$ (see [4]). The formula of Libri and Lebesgue has been generalized by Weil [3]. The material in this note is not new but we hope that perhaps the presentation is.

2. Two properties of $\omega(n)$. For any integer $n \geqq 2$ let $\omega(n) = \exp(2\pi i/n)$. A well-known property possessed by $\omega(n)$ is the following:

LEMMA 1. *If m is an integer, then*

$$\sum_{r=0}^{n-1} \{\omega(n)\}^{mr} = \begin{cases} n, & \text{if } m \equiv 0 \pmod{n}, \\ 0, & \text{if } m \not\equiv 0 \pmod{n}. \end{cases}$$

Proof. The left-hand side is just a geometric progression.

This property of $\omega(n)$ guarantees that any complex-valued function $f(m)$ (m an integer) which is periodic with period n has a finite Fourier series.

LEMMA 2. *If $f(m)$ is periodic in m with period n, then*

$$f(m) = \sum_{r=0}^{n-1} a(r)\{\omega(n)\}^{mr},$$

where $a(r) = (1/n)\sum_{s=0}^{n-1} f(s)\{\omega(n)\}^{-rs}$.

Proof. We have, using Lemma 1,

$$\sum_{r=0}^{n-1} a(r)\{\omega(n)\}^{mr} = \frac{1}{n}\sum_{r=0}^{n-1}\sum_{s=0}^{n-1} f(s)\{\omega(n)\}^{(m-s)r}$$

$$= \frac{1}{n}\sum_{s=0}^{n-1} f(s)\sum_{r=0}^{n-1}\{\omega(n)\}^{(m-s)r} = f(m).$$

3. Counting function. Lemma 1 provides us with a counting function for congruences modulo n, for if $f(x_1, \cdots, x_k)$ is a polynomial with integral coefficients, then the number of solutions $(x_1, \cdots, x_k)$ of $f(x_1, \cdots, x_k) \equiv 0 \pmod{n}$ satisfying $0 \leqq x_i < n$ is given by

$$\sum_{x_1, \cdots, x_k = 0}^{n-1} \left\{ \frac{1}{n} \sum_{r=0}^{n-1} \{\omega(n)\}^{f(x_1, \cdots, x_k) r} \right\} = \frac{1}{n} \sum_{r=0}^{n-1} \sum_{x_1, \cdots, x_k = 0}^{n-1} \{\omega(n)\}^{r f(x_1, \cdots, x_k)}.$$

This can be simplified if $f(x_1, \cdots, x_k)$ is separable in the variables $x_1, \cdots, x_k$. We consider an application where this is so.

4. Application to $x_1^l + \cdots + x_k^l$. We take $f(x_1, \cdots, x_k) = x_1^l + \cdots + x_k^l$ and a prime $n = p \equiv 1 \pmod{l}$, and use the law of exponents: $\omega^{a+b} = \omega^a \omega^b$. Then the number $N_p(l, k)$ of solutions $(x_1, \cdots, x_k)$ of $x_1^l + \cdots + x_k^l \equiv 0 \pmod{p}$ is given by

$$\begin{aligned} N_p(l, k) &= \frac{1}{p} \sum_{r=0}^{p-1} \sum_{x_1, \cdots, x_k = 0}^{p-1} \{\omega(p)\}^{r(x_1^l + \cdots + x_k^l)} \\ &= \frac{1}{p} \sum_{r=0}^{p-1} \left\{ \sum_{x=0}^{p-1} \{\omega(p)\}^{r x^l} \right\}^k . \end{aligned}$$

Let us write $S_p(l, r) = \sum_{x=0}^{p-1} \{\omega(p)\}^{rx^l}$. We note that $S_p(l, r)$ is periodic in r with period p, and

$$S_p(l, 0) = \sum_{x=0}^{p-1} 1 = p,$$

so that $pN_p(l, k) - p^k = \sum_{r=1}^{p-1} \{S_p(l, r)\}^k$. In the summation, r takes on the values $1, 2, \cdots, p-1$. These are $g^0, g^1, g^2, \cdots, g^{p-2}$ (taken modulo p) in some order, where g is a primitive root modulo p. As $S_p(l, r)$ is periodic with period p, we have

$$pN_p(l, k) - p^k = \sum_{s=0}^{p-2} \{S_p(l, g^s)\}^k. \tag{4.1}$$

We next show that $S_p(l, g^s)$ is periodic in s with period l.

LEMMA 3. *For all integers s, $S_p(l, g^s) = S_p(l, g^{s+l})$.*

Proof. We have

$$S_p(l, g^{s+l}) = \sum_{x=0}^{p-1} \{\omega(p)\}^{x^l g^{s+l}} = \sum_{x=0}^{p-1} \{\omega(p)\}^{(gx)^l g^s}.$$

Now the mapping $x \to g^{-1}x$ (so that $gx \to x$) taken modulo p is a bijection on $\{0, 1, \cdots, p-1\}$, so that

$$\sum_{x=0}^{p-1} \{\omega(p)\}^{(gx)^l g^s} = \sum_{x=0}^{p-1} \{\omega(p)\}^{x^l g^s},$$

that is, $S_p(l, g^{s+l}) = S_p(l, g^s)$ as required.

As $p-1 \equiv 0 \pmod{l}$ this periodicity implies

$$\sum_{s=0}^{p-2} \{S_p(l, g^s)\}^k = q \sum_{s=0}^{l-1} \{S_p(l, g^s)\}^k,$$

so that (4.1) becomes

$$pN_p(k, l) - p^k = q \sum_{s=0}^{l-1} \{S_p(l, g^s)\}^k.$$

Now let us examine $S_p(l, g^s)$ (for $s = 0, 1, \cdots, l-1$). We have

$$\begin{aligned} S_p(l, g^s) &= \sum_{x=0}^{p-1} \{\omega(p)\}^{g^s x^l} = 1 + \sum_{x=1}^{p-1} \{\omega(p)\}^{g^s x^l} \\ &= 1 + \sum_{t=0}^{p-2} \{\omega(p)\}^{g^{s+lt}} = 1 + \sum_{r=0}^{l-1} \sum_{u=0}^{q-1} \{\omega(p)\}^{g^{s+l(qr+u)}}. \end{aligned}$$

But $g^{p-1} \equiv 1$, so we have

$$S_p(l, g^s) = 1 + \sum_{r=0}^{l-1} \sum_{u=0}^{q-1} \{\omega(p)\}^{g^{s+lu}} = 1 + l \sum_{u=0}^{q-1} \{\omega(p)\}^{g^{s+lu}}.$$

The expressions $\sum_{u=0}^{q-1} \{\omega(p)\}^{g^{s+lu}}$ are called the *q-nomial periods* of the pth roots of unity [2]. We write

$$\eta_s = \sum_{u=0}^{q-1} \{\omega(p)\}^{g^{s+lu}}$$

so that we have the result of Libri and Lebesgue [1]:

THEOREM 1. *The number $N_p(k, l)$ of solutions of $x_1^l + \cdots + x_k^l \equiv 0 \pmod{p}$ is given by $N_p(k, l) = p^{k-1} + (q/p) \sum_{s=0}^{l-1} \{1 + l\eta_s\}^k$.*

5. An alternative expression for $N_p(k, l)$. We can apply Lemma 2 to $S_p(l, g^s)$ (as it is periodic in s with period l) to obtain a different expression for $S_p(l, g^s)$ and thus a different expression for $N_p(k, l)$. By Lemma 2 we have $S_p(l, g^s) = \sum_{r=0}^{l-1} a(r) \{\omega(l)\}^{rs}$, where

$$\begin{aligned} a(r) &= \frac{1}{l} \sum_{s=0}^{l-1} S_p(l, g^s) \{\omega(l)\}^{-rs} \\ &= \frac{1}{l} \sum_{s=0}^{l-1} \left\{1 + l \sum_{u=0}^{q-1} \{\omega(p)\}^{g^{s+lu}}\right\} \{\omega(l)\}^{-rs} \\ &= \frac{1}{l} \sum_{s=0}^{l-1} \{\omega(l)\}^{-rs} + \sum_{s=0}^{l-1} \sum_{u=0}^{q-1} \{\omega(p)\}^{g^{s+lu}} \{\omega(l)\}^{-rs} \end{aligned}$$

$$= \frac{1}{l}\sum_{s=0}^{l-1}\{\omega(l)\}^{-rs} + \sum_{t=0}^{p-2}\{\omega(p)\}^{g^t}\{\omega(l)\}^{-rt}$$

$$= \begin{cases} 1 + \sum_{t=0}^{p-2}\{\omega(p)\}^{g^t}, & r = 0, \\ 0 + \sum_{t=0}^{p-2}\{\omega(p)\}^{g^t}\{\omega(l)\}^{-rt}, & r = 1, 2, \cdots, l-1, \end{cases}$$

$$= \begin{cases} 0, & r = 0, \\ \sum_{t=0}^{p-2}\{\omega(p)\}^{g^t}\{\omega(l)\}^{-rt}, & r \neq 0. \end{cases}$$

Writing $\tau_r = \sum_{t=0}^{p-2}\{\omega(p)\}^{g^t}\{\omega(l)\}^{-rt}$, where r is any integer, we have

$$S_p(l, g^s) = \sum_{r=1}^{l-1}\{\omega(l)\}^{sr}\tau_r,$$

giving the following theorem:

THEOREM 2. *The number $N_p(k, l)$ of solutions of $x_1^l + \cdots + x_k^l \equiv 0 \pmod p$ is given by*

$$N_p(k, l) = p^{k-1} + \frac{q}{p}\sum_{s=0}^{l-1}\left\{\sum_{r=1}^{l-1}\{\omega(l)\}^{sr}\tau_r\right\}^k.$$

This generalizes a result of the author [4] when $l=3$, viz.,

$$N_p(k, 3) = p^{k-1} + [(p-1)/3p][(\tau_1+\tau_2)^k + (\omega\tau_1+\omega^2\tau_2)^k + (\omega^2\tau_1+\omega\tau_2)^k],$$

where $\omega \equiv \omega(3) = \frac{1}{2}(-1+\sqrt{-3})$.

References

1. V. A. Lebesgue, Recherches sur les nombres, J. Math. Pures Appl., 2 (1837) 253–292.
2. G. B. Mathews, Theory of Numbers, Chelsea, New York, 1962, p. 191.
3. A. Weil, Numbers of solutions of equations in finite fields, Bull. Amer. Math. Soc., 55 (1949) 497–508.
4. K. S. Williams, On the number of solutions of a congruence, this MONTHLY, 73 (1966) 44–49.

LIMIT POINTS OF SEQUENCES IN METRIC SPACES

M. D. AŠIĆ AND D. D. ADAMOVIĆ, University of Belgrade, Yugoslavia

The aim of this paper is to generalize both statements of the following theorem [1]:

THEOREM A. *Let $C(\xi)$ be the set of limit points of the bounded complex sequence ξ. Then $C(\xi)$ is connected if and only if there exists a subsequence $\eta = (y_n)$ of ξ such that $C(\eta) = C(\xi)$ and $y_{n+1} - y_n \to 0$ $(n \to \infty)$.*

The "if part" of this theorem is due to H. G. Barone [2], and the "only if part" to Paul Schaeffer [1].

A generalization of the first statement is

THEOREM 1. *Let X be a compact metric space, $\xi=(x_n)$ a sequence of its points, and $C(\xi)$ the set of limit points of ξ. Assume*

$$\lim_{n\to\infty} d(x_n, x_{n+1}) = 0. \tag{1}$$

Then $C(\xi)$ is a connected set.

Proof. Suppose that $C(\xi)$ is not connected, i.e., that

$$C(\xi) = A \cup B \quad \text{and} \quad A \cap B = \varnothing, \tag{2}$$

with A and B nonempty closed sets. Put

$$\epsilon = \operatorname{Inf}\{d(t, u) : t \in A, u \in B\}. \tag{3}$$

Since the space X is compact, there exist $a\in A$ and $b\in B$ such that $d(a, b)=\epsilon$. In view of the second condition of (2), we have $\epsilon>0$.

The sets

$$O_1 = \bigcup_{t\in A} K(t, \epsilon/3) \quad \text{and} \quad O_2 = \bigcup_{t\in B} K(t, \epsilon/3), \tag{4}$$

where $K(t, \delta)$ is the open sphere whose center is $t\in X$ and radius is δ, are then disjoint. Owing to the compactness of X, the set $O_1\cup O_2$ contains all members of the sequence ξ with large enough indices. This last fact together with (1) implies that there is a natural number n_0 such that

$$x_n \in O_1 \cup O_2 \qquad (n \geqq n_0) \tag{5}$$

and

$$d(x_n, x_{n+1}) < \epsilon/3 \qquad (n \geqq n_0). \tag{6}$$

Since A and B are nonempty sets of limit points of the sequence ξ, upon taking into account (4) and (5), we see there exists a natural number $m\geqq n_0$, such that $x_m\in O_1$ and $x_{m+1}\in O_2$, and so there exist $t\in A$ and $u\in B$, such that

$$d(t, x_m) < \epsilon/3, \quad d(u, x_{m+1}) < \epsilon/3. \tag{7}$$

Then, by (6) and (7),

$$d(t, u) \leqq d(t, x_m) + d(x_m, x_{m+1}) + d(x_{m+1}, u) < \epsilon,$$

which contradicts the choice of ϵ, and thus proves the theorem.

For the proof of our second theorem, which generalizes the second statement of Theorem A, we shall need the following lemma (see [3], pp. 169–170, T, (a), where the supposition of compactness of X is not necessary).

LEMMA. *Let E be a connected part of a metric space. Then for any two elements t and u of E and for every $\epsilon>0$ there is a finite number of elements $t_0, t_1, \cdots, t_k$ of E, such that $t_0=t$, $t_k=u$ and $d(t_{\nu-1}, t_\nu)<\epsilon$ $(\nu=1, \cdots, k)$.*

THEOREM 2. *Let X be a metric space and $\xi=(x_n)$ a sequence of its points, for which $C(\xi)$ is non-empty and connected. Then there exists a subsequence $\eta=(y_n)$ of the sequence ξ with the properties $\lim_{n\to\infty} d(y_n, y_{n+1})=0$ and $C(\eta)=C(\xi)$.*

Proof. For each $\delta>0$, there exists an (at most) countable cover Σ of the set $C(\xi)$ which consists of spheres of diameter δ whose intersections with $C(\xi)$ are not empty. Namely, Σ can be formed from all those spheres with centers in x_n $(n=1, 2, 3, \cdots)$ and with diameter δ which have nonempty intersections with $C(\xi)$. If Σ_ν denotes such a collection of spheres for $\delta=\nu^{-1}$ $(\nu=1, 2, 3, \cdots)$, the collection of spheres $\Sigma_0=\bigcup_{\nu=1}^{\infty}\Sigma_\nu$ is countable, so that we can write $\Sigma_0=\{K_1, K_2, K_3, \cdots\}$. Besides, for each t in $C(\xi)$ there exists a sphere from Σ_0 with an arbitrary small diameter which contains t. Let $t_i\in K_i\cap C(\xi)$ $(i=1, 2, 3, \cdots)$. According to the lemma, there exist positive integers $1=k_1<k_2<\cdots<k_i<\cdots$ and points $u_n\in C(\xi)$ $(n=1, 2, 3, \cdots)$ such that

$$u_{k_i} = t_i (i = 1, 2, 3, \cdots);\ d(u_{n-1}, u_n) < i^{-1}(k_{i-1} < n \leqq k_i;\ i = 2, 3, \cdots).$$

It is clear that

$$\lim_{n\to\infty} d(u_n, u_{n+1}) = 0. \tag{8}$$

Furthermore, let p_1 be the smallest positive integer such that $x_{p_1}\in K_1$ and $d(x_{p_1}, u_1)<1$; if the natural numbers p_ν $(\nu=1, \cdots, n-1)$ are determined, then let p_n be the smallest natural number with the properties:

$$p_n > p_{n-1};\ d(x_{p_n}, u_n) < n^{-1};\quad \text{if } n = k_i, \text{ then } x_{p_n} \in K_i.$$

According to the construction of the sequence (u_n), an infinite and strictly monotonic sequence of natural numbers (p_n) has been defined by the above procedure. The sequence $\eta=(y_n)$, where $y_n=x_{p_n}$ $(n=1, 2, 3, \cdots)$, has, therefore, members with arbitrary large indices in each sphere K_i $(i=1, 2, 3, \cdots)$. Therefore, $C(\eta)\supset C(\xi)$, and since the opposite inclusion clearly holds, we have $C(\eta)=C(\xi)$. On the other hand, according to (8),

$$\begin{aligned} d(y_n, y_{n+1}) &= d(x_{p_n}, x_{p_{n+1}}) \leqq d(x_{p_n}, u_n) + d(u_n, u_{n+1}) + d(u_{n+1}, x_{p_{n+1}}) \\ &< n^{-1} + d(u_n, u_{n+1}) + (n+1)^{-1} \to 0 \qquad (n\to\infty). \end{aligned}$$

A few additional remarks concerning Theorem 1.

1. Compactness of the space X cannot be simply omitted from Theorem 1. This can be proved by choosing a suitable sequence ξ in the two-dimensional Euclidean space $X=R^2$ (for example, such that $C(\xi)$ be the union of a curve and its asymptote).

2. On the other hand, compactness of the space X is not necessary for the

holding of the implication

$$\lim_{n\to\infty} d(x_n,\, x_{n+1}) = 0 \Rightarrow C(\xi) \text{ is a connected set,} \tag{9}$$

which is proved by the example of the one-dimensional Euclidean space $X=R$.

3. Another example for the same statement is a metric space X with a nondenumerable amount of points in which $d(x, y)=1$ $(x\neq y)$.

4. In view of the above, we can raise the question of replacing the compactness of X in Theorem 1 by a weaker condition. Such a condition certainly could not be Σ-compactness (the property of a space to be represented as a denumerable union of compact parts). Moreover, according to Remarks 1 and 3, Σ-compactness of the space X is neither a necessary nor sufficient condition for implication (9) to hold.

References

1. Paul Schaefer, Limit points of bounded sequences, this MONTHLY, 75 (1968) 51.
2. H. G. Barone, Limit points of sequences and their transforms by methods of summability, Duke Math. J., 5 (1939) 740–752.
3. J. L. Kelley, General Topology, Van Nostrand, Princeton, N. J., 1961.

RESEARCH PROBLEMS

EDITED BY VICTOR KLEE

In this Department the Monthly presents easily stated research problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Material should be sent to Victor Klee, Department of Mathematics, University of Washington, Seattle, WA 98105.

MUST A COMPACT ENDSET HAVE MEASURE ZERO?

VICTOR KLEE AND MICHAEL MARTIN, University of Washington

Statement of the problem. For any collection $\mathcal{S}$ of line segments in Euclidean d-space E^d, let $P(\mathcal{S})$ denote the set of all endpoints of the members of $\mathcal{S}$. Any set of two or more points has the form $P(\mathcal{S})$ for some $\mathcal{S}$, but a subset M of E^d is here called an *endset* if and only if $M=P(\mathcal{S})$ for some collection $\mathcal{S}$ of pairwise disjoint segments. In the real line E^1 any endset is countable. Intuition suggests that the d-dimensional Lebesgue measure of any compact endset in E^d is zero. When $d\leqq 2$ this follows from an argument in [**3**, 3.3]. However, for $d=4$ (and hence also for all $d>4$) Bruckner and Ceder [**1**] have produced a counterexample by using Nikodym's construction [**5**] of a Cantor set X of positive measure in E^2 such that for each point x of X there is a line in E^2 intersecting X only at

x. Hence the answer to the title question is affirmative for $d \leqq 2$, unknown for $d=3$, and negative for $d \geqq 4$.

Origin of the problem. The problem originated in a study [3] of the facial structure of convex bodies. For any convex set C in E^d let $I(C)$ denote the interior of C relative to the smallest flat containing C. Now suppose that B is a convex body in E^d, X is the boundary of B, and X_u is the union of the sets $I(C)$ as C ranges over all maximal convex subsets of X. For example, if B is a polytope X_u consists of the entire boundary except for points on $(d-2)$-dimensional faces of B. It seems probable that, as surely happens when B is a polytope, X_u is almost all of X in the sense that the $(d-1)$-dimensional measure of the set $X \sim X_u$ is equal to zero. For $d \leqq 3$ this is proved in [3, 3.4] by using the fact that in E^1 and E^2, compact endsets have measure zero. It may be that the conjecture concerning convex bodies is correct for all d even though (in view of the Bruckner-Ceder example) this method of proof is inapplicable for $d \geqq 5$.

An equivalent question. There is an endset M_f in E^d associated in a natural way with any real-valued function f whose domain is a subset D_f of E^{d-1}. The endset M_f is the union of the graph of f and the graph of the function $f+1$; that is,

$$M_f = \{(x, f(x)) : x \in D_f\} \cup \{(x, f(x) + 1) : x \in D_f\} \subset E^{d-1} \times E^1 = E^d.$$

As the various sets $M_{f+\tau}$, for $0<\tau<1$, are pairwise disjoint and all translation-equivalent to M_f, it follows that either M_f is nonmeasurable (a possibility that can be realized in various ways—see, for example, [2, 21.27]) or the d-dimensional Lebesgue measure of M_f is zero. A similar argument shows the measure of $M=P(\mathcal{S})$ is zero whenever M is measurable and $\mathcal{S}$ is a collection of pairwise disjoint *parallel* segments. As is explained in the next paragraph, the title question amounts to asking whether *almost parallel* is as good as parallel in this context.

For each number $\eta \in]0, 1/3[$ define a (d, η)-*endset* as a compact set of the form $P(\mathcal{S})$, where $\mathcal{S}$ is a collection of pairwise disjoint segments having one endpoint within η of the origin $(0, \cdots, 0, 0)$ in E^d and the other endpoint within η of the point $(1, 0, \cdots, 0)$. The segments in such a collection need not be parallel, but they are very nearly so, especially when η is close to zero (see Figure 1).

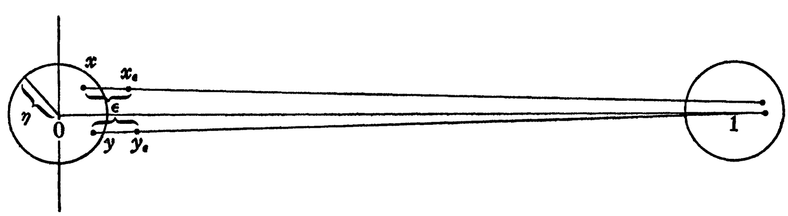


FIG. 1

Using standard and elementary techniques of measure theory, it is easy to derive

(a) any measurable endset in E^d is of measure zero

from

(b) there exists $\eta_d > 0$ such that any (d, η_d)-endset is of measure zero.

For $d=2$ the truth of (b) follows from an argument in [3, 3.3]. The proof depends on the fact that if $P(\mathcal{S})$ is a $(2, \eta)$-endset for small enough η, and x and y are the left endpoints of two members of $\mathcal{S}$, then

$$\|x_\epsilon - y_\epsilon\| \geqq \|x - y\|/2$$

for all $\epsilon \in]0, \eta[$, where x_ϵ and y_ϵ are obtained from x and y by moving these points a distance ϵ toward the corresponding right endpoints (see Figure 1). It follows that the measure of the set of all left endpoints is not much reduced by the ϵ-motion, and an argument similar to the one above (involving an uncountable collection of pairwise disjoint sets) can be applied to show that $P(\mathcal{S})$ has measure zero. However, this doesn't work when $d \geqq 3$, for then the two segments may nearly cross and $\|x_\epsilon - y_\epsilon\|$ may be much smaller than $\|x-y\|$ (Figure 2).

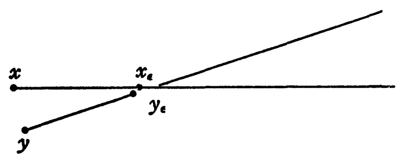


FIG. 2

An example and a problem in E^2. Let X be a Cantor set in E^1 such that the 1-dimensional measure of X is positive. Then of course the 2-dimensional measure of $X \times X$ is positive. Further, $X \times X$ is contained in a simple closed curve J [4, p. 540] and there is a homeomorphism of E^2 onto itself which carries J onto an ordinary circle K [4, p. 173]. As K is plainly the set of all endpoints of a collection of pairwise disjoint topological arcs, the same is true of J, even though the 2-dimensional measure of J is positive. It would be of interest, for a collection $\mathcal{S}$ of pairwise disjoint topological arcs, to find geometric conditions that are weaker than straightness and yet guarantee that the set of all endpoints has measure zero if it is measurable.

References

1. A. M. Bruckner and J. Ceder, A note on end-sets, to appear.

2. E. Hewitt and K. Stromberg, Real and Abstract Analysis, Springer, Berlin, 1965.

3. V. Klee and M. Martin, Semicontinuity of the face-function of a convex set, Comment. Math. Helv., to appear.

4. K. Kuratowski, Topology, vol. II, Academic Press, New York, 1968.

5. O. M. Nikodym, Sur la mesure des ensembles plans dont tous les points sont rectilinéairement accessibles, Fund. Math., 10 (1927) 116–168.

SOME PROBLEMS IN ADDITIVE NUMBER THEORY

P. ERDÖS, Israel Institute of Technology, Haifa

1. Let $1 \leqq a_1 < \cdots < a_k \leqq x$ be a sequence of integers for which the sums

$$\sum_{i=1}^{k} \epsilon_i a_i, \qquad \epsilon_i = 0 \quad \text{or} \quad 1 \tag{1}$$

are all distinct. Put max $k = f(x)$ where the maximum is taken over all sequences satisfying (1). It is easy to see that

$$f(x) < \frac{\log x}{\log 2} + \frac{\log \log x}{\log 2} + O(1),$$

and Moser and I proved

$$f(x) < \frac{\log x}{\log 2} + \frac{\log \log x}{2 \log 2} + O(1). \tag{2}$$

This is the best-known upper bound for $f(x)$.

Is it true that

$$f(x) = (\log x/\log 2) + O(1)? \tag{3}$$

Moser and I asked: Is it true that $f(2^k) \geqq k+2$ for sufficiently large k? Conway and Guy showed that the answer is affirmative (unpublished).

P. Erdös, Problems and results in additive number theory, *Colloque, Théorie des Nombres*, Bruxelles 1955, p. 137.

2. Let $1 \leqq a_1 < \cdots < a_k \leqq x$ be a sequence of integers so that all the sums

$$a_{i_1} + \cdots + a_{i_s}, \quad i_1 \leqq i_2 \leqq \cdots \leqq i_s, \qquad 1 \leqq s \leqq r$$

are distinct. Put max $k = g_r(x)$. Turán and I proved

$$g_2(x) < x^{1/2} + O(x^{1/4}). \tag{4}$$

This was recently improved by Lindström to $g_2(x) \leqq x^{1/2} + x^{1/4} + 1$. The lower bound $g_2(x) \geqq (1+o(1))x^{1/2}$ easily follows from a classical result of Singer on difference sets. Turán and I conjectured

$$g_2(x) = x^{1/2} + O(1). \tag{5}$$

Bose and Chowla proved that $g_r(x) \geqq (1+o(1))x^{1/r}$ for each $r \geqq 2$, and they conjectured

$$g_r(x) = (1 + o(1))x^{1/r}. \tag{6}$$

This is known only for $r = 2$.

R. C. Bose and S. Chowla, *Theorems in the additive theory of numbers*, Comment. Math. Helv., 37 (1962–63) 141–147.

H. Halberstam and K. Roth, *Sequences*, Oxford Univ. Press, Oxford 1966.

P. Erdös and P. Turán, *On the Problem of Sidon in additive number theory and on some related problems*, Journal London Math. Soc., 16 (1941) 212–215 and 19 (1944) 208.

J. Singer, *A theorem in finite projective geometry and some applications to number theory*, Trans. Amer. Math. Soc., 43 (1938) 377–385.

B. Lindström, *An inequality for B_2 sequences*, Journal Combinatorial Theory, 6 (1969) 211–212.

3. Let $1 \leqq a_1 < \cdots$ be an infinite sequence of integers. Denote by $h(n)$ the number of solutions of $n = a_i + a_j$. Turán and I conjectured that if $h(n) > 0$ for $n > n_0$ then

$$\limsup_{n\to\infty} h(n) = \infty. \tag{7}$$

Another stronger conjecture states that if $a_k < ck^2$ for every k then (7) holds.

These conjectures seem very difficult. It is a curious fact that the multiplicative analogues of these conjectures are not intractable.

Let $1 < b_1 < \cdots$ be an infinite sequence of integers. Denote by $H(n)$ the number of solutions of $n = b_i b_j$. Assume $H(n) > 0$ for $n > n_0$. I proved

$$\limsup_{n\to\infty} H(n) = \infty.$$

Turán and I further conjectured that the number of solutions of $a_i + a_j \leqq x$ cannot be of the form $cx + O(1)$, where $c < \infty$. In other words

$$\sum_{n=1}^{x} h(n) = cx + O(1) \tag{8}$$

can hold only if $c = 0$ and the sequence, $a_1 < \cdots$, is finite. Fuchs and I proved a very much stronger result than (8). We in fact showed that if $c > 0$ then

$$\sum_{n=1}^{x} h(n) = cx + o\left(\frac{x^{1/4}}{(\log x)^{1/2}}\right) \tag{9}$$

is impossible. Jurkat showed (unpublished) that (9) is impossible even with $o(x^{1/4})$. Perhaps Jurkat's result is best possible and

$$\sum_{n=1}^{x} h(n) = cx + O(x^{1/4})$$

can hold for a suitable sequence $a_1 < \cdots$.

Is it true that the number of solutions of $a_1 + a_j + a_r \leqq x$ cannot be of the form $cx + O(1)$? Our method used with Fuchs does not apply here.

I have several times offered \$250 for the proof or disproof of any of the conjectures (3), (5), or (7).

Paul Erdös, *On extremal problems of graphs and generalised graphs*, Israel Journal of Math., 2 (1965) 183–190, and *On the multiplicative representation of*

integers, *ibid.*, 2 (1965) 251–261.

Paul Erdös and W. H. J. Fuchs, *On a problem of additive number theory*, Journal London Math. Soc., 31 (1956) 67–73. See also the well-known book: H. Halberstam and K. Roth, *Sequences*, Oxford Univ. Press, Oxford, 1966.

A. Stohr, *Gelöste und ungelöste Fragen über Basen der natürlichen Zahlenreihe*, I and II, Journal reine u. angew. Math., 194 (1955) 40–65 and 111–140.

CLASSROOM NOTES

EDITED BY DAVID DRASIN

Manuscripts for this Department should be sent to David Drasin, Division of Mathematical Sciences, Purdue University, Lafayette, IN 47907.

AN EXISTENCE THEOREM FOR NON-NOETHERIAN RINGS

ROBERT GILMER, Florida State University

In developing the theory of Noetherian rings, it is desirable to have at hand some examples of non-Noetherian rings. One such example is the ring of polynomials in infinitely many indeterminates over any nonzero ring. A second example is $R[X]$, where R is any nonzero commutative ring without identity [1], but in this case, $R[X]$ is also a ring without identity. We present here a theorem which provides a method for constructing, as subrings of a polynomial ring in finitely many indeterminates, a wide class of commutative rings with identity which are not Noetherian. It should be noted that the proof of the theorem given uses only one result (Lemma 1) outside the basic theory of Noetherian rings, namely, that a finitely generated idempotent ideal of a commutative ring is principal and is generated by an idempotent element. An examination of the proof of this result reveals that even it is obtained as a direct application of Cramer's Rule for determinants over a commutative ring with identity.

THEOREM 1. *Suppose that R is a nonzero commutative ring, that A is a nonzero ideal of R distinct from R, and that $\{X_\lambda\}_{\lambda\in\Lambda}$ is a set of indeterminates over R. The subring $S=R+A[\{X_\lambda\}]$ of $R[\{X_\lambda\}]$, consisting of those polynomials over R having each of their nonconstant coefficients in A, is Noetherian if and only if these three conditions hold*: (1) *Λ is finite*, (2) *R is Noetherian, and* (3) *the ideal A is idempotent.*

Proof. Suppose that S is Noetherian. It is clear that Λ must be finite, for if not, the ideal $A[\{X_\lambda\}]$ of S would not be finitely generated. The mapping on S which sends each polynomial $f\in S$ onto its constant term is a homomorphism from S onto R, so that R is Noetherian. If σ is a fixed element of Λ and if $a\in A$, the ideal $(aX_\sigma, aX_\sigma^2, \cdots, aX_\sigma^m, \cdots)$ of S is finitely generated, so that $aX_\sigma^{m+1} \in (aX_\sigma, \cdots, aX_\sigma^m)$ for some positive integer m:

$$aX_\sigma^{m+1} = \sum_{i=1}^{m} (s_i a X_\sigma^i + n_i a X_\sigma^i)$$

for some $s_i \in S$, $n_i \in Z$. Equating coefficients of X_σ^{m+1} from each side of this equality, we conclude that $a = ba$ for some element b in A. Thus $a \in A^2$, $A \subseteq A^2$, and therefore $A = A^2$.

To prove the other half of Theorem 1, we need to prove a lemma. This result is known (for example, a proof appears in [2], p. 58); we include its proof here for the sake of completeness.

LEMMA 1. *If B is a finitely generated idempotent ideal of a commutative ring T, then B is principal and is generated by an idempotent element.*

Proof. We first assume that T has an identity and we let $\{b_i\}_{i=1}^n$ be a finite set of generators for B. Then $B = B^2 = \sum_{i=1}^n Bb_i$ so that we obtain a system of equations

$$b_k = \sum_{i=1}^{n} s_{ki} b_i,$$

where $s_{ki} \in B$ and where $1 \leqq k \leqq n$. This gives rise to the system of equations

$$\sum_{i=1}^{n} (\delta_{ki} - s_{ki}) b_i = 0 \qquad 1 \leqq k \leqq n,$$

where δ_{ki} is the Kronecker delta.

By Cramer's rule, $db_i = 0$ for $1 \leqq i \leqq n$, where d is the determinant of the matrix $[\delta_{ki} - s_{ki}]$. It is easy to see, however, that d is of the form $1 - b$ for some element b in B. Since $0 = db_i = b_i - bb_i$ for each i, B is the principal ideal generated by b. And since $1 - b$ annihilates B, $(1-b)b = 0$, or $b = b^2$.

If T contains no identity element, we consider a commutative ring T^* obtained by adjoining an identity element to T. Then B is a finitely generated idempotent ideal of T^*, and hence is principal as an ideal of T^*, generated by an idempotent element v. Since T^* is obtained by adjoining an identity element to T, it follows that v also generates B as an ideal of T.

To complete the proof of Theorem 1, we assume that Λ is finite—say $\Lambda = \{1, 2, \cdots, t\}$, that R is Noetherian, and that A is idempotent. Then A is principal and is generated by an idempotent element e. Moreover, R is the direct sum of A and the ideal $B = \{x - ex \mid x \in R\}$. Since R is Noetherian, $A \simeq R/B$ and $B \simeq R/A$ are Noetherian rings, and A is a ring with identity. Hence, $A[X_1, \cdots, X_t]$ and B are ideals of S which are Noetherian rings. It then follows that

$$S = R + A[X_1, \cdots, X_t] = B \oplus A[X_1, \cdots, X_t]$$

is a Noetherian ring.

COROLLARY. *Suppose that R is a commutative ring with identity containing no idempotent elements other than* 0 *and* 1. *If A is a nonzero ideal of R distinct from R,*

and if $\{X_\lambda\}_{\lambda\in\Lambda}$ *is any set of indeterminates over* R, *then* $S=R+A[\{X_\lambda\}]$ *is a commutative ring with identity which is not Noetherian.*

References

1. R. Gilmer, If $R[X]$ is Noetherian, R contains an identity, this MONTHLY, 74 (1967) 700.
2. ———, Multiplicative Ideal Theory, Queen's University, Kingston, Ontario, 1968.

MATHEMATICAL EDUCATION

EDITED BY J. G. HARVEY AND M. W. POWNALL

Material for this Department should be sent to either of the editors: J. G. Harvey, Department of Mathematics, University of Wisconsin, Madison, WI 53706; M. W. Pownall, Department of Mathematics, Colgate University, Hamilton, NY 13346.

THE COSRIMS REPORTS

R. P. BOAS, Northwestern University

What is—or was—Cosrims? It was a 12-man committee of the National Academy of Sciences under the chairmanship of Lipman Bers, and the name stands for Committee on the Support of Research in the Mathematical Sciences. Before it was through it had 50 or more collaborators doing things for it, and another 50 or so worked on the CBMS Survey, with John Jewett as executive director, that collected the needed data. The principal product of Cosrims was simply entitled "The Mathematical Sciences—a Report"; it was completed, after more than a year of work, at the end of 1967 and issued late in 1968.

You might well wonder what Cosrims is doing on this program, when it is AMS rather than MAA that is primarily concerned with research in mathematics. The answer is that one of Cosrims' main principles, from the very beginning, was that one cannot separate research and education in mathematics. This is because the increasing mathematization of society, which we cannot stop even if we want to, demands that more and more people should be trained to some degree in mathematics (and when I say "mathematics" I mean the mathematical sciences generally); the people who train them must themselves be trained at a higher level, and this training must eventually be done by the relatively small body of mathematical scientists who do research—an activity that itself calls for continued learning. Even the purest mathematician I know, one who says flatly that he is interested *only* in research in pure mathematics, at least adds, "and in training other people to do research in pure mathematics."

Let me quote from the Cosrims report:

We need people who can teach mathematics in grade school in a way that will not create a permanent psychological block against mathematics in so many of our fellow citizens. We need people who can understand a simple formula, read a graph, interpret a statement about prob-

PROBLEMS AND SOLUTIONS

EDITED BY EMORY P. STARKE

All problems (both elementary and advanced) proposed for inclusion in this Department should be sent to E. P. Starke, 1000 Kensington Ave., Plainfield, NJ 07060. Proposers of problems are urged to enclose any solutions or information that will assist the editors. Ordinarily, problems in well-known textbooks and results in generally accessible sources are not appropriate for this Department. No solution (except those accompanying proposals) should be sent to Professor Starke.

ELEMENTARY PROBLEMS

Solutions of Elementary Problems should be sent to Problems Group, Mathematics Department, University of Maine, Orono, ME 04473. To facilitate their consideration, solutions of Elementary Problems in this issue should be typed (with double spacing) and should be mailed before September 30, 1970. Contributors (in the United States) who desire acknowledgment of receipt of their solutions are asked to enclose self-addressed stamped postcards.

E 2242. *Proposed by M. L. Fredman, California Institute of Technology.*

Show that $\sum_{i=1}^{n} k^{(i,n)} \equiv 0 \pmod{n}$, $n \geqq 1$, for all integers k (positive or negative), where (a, b) denotes the greatest common divisor.

E 2243. *Proposed by Richard Sinkhorn, University of Houston.*

Show that every normal stochastic matrix is necessarily doubly stochastic.

E 2244. *Proposed by T. K. Leong, T. A. Peng and K. C. Yeo, University of Singapore*

Show that, for any fixed $m \geqq 2$, the series

$$1 + \frac{1}{2} + \cdots + \frac{1}{m-1} - \frac{x}{m} + \frac{1}{m+1} + \frac{1}{m+2} + \cdots + \frac{1}{2m-1} - \frac{x}{2m}$$
$$+ \frac{1}{2m+1} + \frac{1}{2m+2} + \cdots + \frac{1}{3m-1} - \frac{x}{3m} + \cdots$$

is convergent for exactly one value of x and find the sum of the series for this x.

E 2245. *Proposed by A. W. Walker, Toronto, Canada*

If A, B, C; a, b, c; s are the angles, side lengths, and semiperimeter of any plane triangle, then

$$(a + b + c)^3(s - a)(s - b)(s - c) \geqq (a^2 + b^2 + c^2)^3 \cos A \cos B \cos C.$$

E 2246. *Proposed by Andrew Rochman, Saint Louis University*

Let f be a non-constant, real valued, continuous function such that, for all x, $y \in R$, $f(x+y)=\phi(f(x), y)$. Prove that f is monotonic.

SOLUTIONS OF ELEMENTARY PROBLEMS

Fixed Point for a Composite Function

E 2187 [1969, 826]. *Proposed by R. E. Chandler, North Carolina State University*

Let D be the closed unit disk in the complex plane and let f, $g:D\to D$. Suppose g is analytic on some open set containing D and suppose that f and its iterates $f^2, f^3, \cdots$ each have exactly one fixed point on D. If f and g commute $(f(g(z))=g(f(z))$ for all $z\in D)$ then either g has exactly one fixed point in D or $g(z)\equiv z$.

I. *Solution by J. R. Kuttler, The Johns Hopkins University Applied Physics Laboratory.* The result is not correct. As a counterexample, let g be analytic as above, not the identity, and have more than one fixed point on D. Let z_0 be a fixed point of g and let $f(z)=z_0$ for all $z\in D$, i.e., f takes all of D onto the point z_0. Then $g(f(z))=g(z_0)=z_0=f(g(z))$ for all $z\in D$, and f and its iterates have the unique fixed point z_0.

II. *Solution by O. P. Lossers, Technological University, Eindhoven, Netherlands.* With the added hypothesis that f be one-one, the result is valid. Let z_0 be the unique fixed point of f. Then z_0 is also the unique fixed point of f^k $(k=2, 3, \cdots)$. Since $f(g(z_0))=g(f(z_0))=g(z_0)$ it follows that z_0 is also a fixed point of g. Now let z_1 be a fixed point of g. Then we have $g(f(z_1))=f(g(z_1))=f(z_1)$. Hence $f(z_1)$ is also a fixed point of g. It follows by induction that $\{z_1, z_2, \cdots\}$ defined by $z_{n+1}=f^n(z_1)$ is a sequence of fixed points of g.

If all the z_k are different, then the function $g(z)-z$ has infinitely many zeros on D. Since g is analytic it follows that $g(z)\equiv z$ in this case. If $g(z)\not\equiv z$ there are positive integers j, k with $j<k$ such that $z_{j+1}=z_{k+1}$. Hence $f^j(z_1)=f^k(z_1)=f^j(f^{k-j}(z_1))$. It follows from the fact that f (and hence f^k) is one-to-one that $z_1=f^{k-j}(z_1)$ and since z_0 is the only fixed point of f^{k-j} we see that $z_0=z_1$. Hence z_0 is the only fixed point of g.

Also solved by M. G. Greening (Australia), M. S. Klamkin, Renate McLaughlin, Norman Miller, Simeon Reich (Israel), Steven Rohde, E. F. Schmeichel, and the proposer.

Series Converging for all Subsets of a Family

E 2188 [1969, 826]. *Proposed by P. R. Chernoff, University of California at Berkeley, and W. C. Waterhouse, Cornell University*

Let $\sum_S a_n$, where S is a subset of the positive integers, mean the sum with elements of S taken in increasing order. Let F be an arbitrary countable family of subsets; construct a conditionally convergent series $\sum a_n$ with $\sum_S a_n$ con-

verging for all S in F. (Several special cases, notably F = all arithmetic progressions and F = all geometric progressions, are problems in Pólya & Szegö, *Aufgaben und Lehrsätze*, I, 3, 3.)

Solution by the proposers. Let $S_1, S_2, \cdots$ be the sets in F. Let T_1 be S_1 if S_1 is infinite; otherwise let T_1 be its complement S_1'. Let T_2 be $T_1 \cap S_2$ if this is infinite; otherwise let $T_2 = T_1 \cap S_2'$. Proceeding inductively we thus construct a sequence of infinite sets $T_1 \supseteq T_2 \supseteq T_3 \supseteq \cdots$ with T_m contained in S_m or S_m'. Choose $t_1 \in T_1$, and inductively let t_{m+1} be the least element of T_{m+1} greater than t_m. Then the t_k for $k \geqq m$ lie all in S_m or all in its complement. Now just set $a_{t_k} = (-1)^k/k$, with all other $a_n = 0$.

Also solved by G. A. Heuer, and Joel Spencer.

An Identity on Restricted Partitions

E 2189 [1969, 826]. *Proposed by S. W. Golomb, University of Southern California*

Let $b(n, k)$ be the number of partitions of the positive integer n into k integer parts all of which are powers of 2. (We allow parts to be repeated; we allow $2^0 = 1$ as a part; and we disregard the order in which the k parts are listed.) Show that

$$1 + \sum_{n=1}^{\infty} \sum_{k=1}^{n} b(n, k) x^n y^k = 1 \Big/ \sum_{n=0}^{\infty} (-y)^{w(n)} x^n,$$

where $w(n)$ is the number of 1's in the binary representation of n. Also, reach a conclusion about these partitions based on the substitution $y = -1$.

Solution by Michael Goldberg, Washington, D. C. It was noted by Euler that the coefficient of $x^n y^k$ in the expansion of

$$F(x, y) = (1 + yx)(1 + yx^2)(1 + yx^4) \cdots \tag{1}$$

is the number of different ways that n is the sum of k different powers of 2. Hence, $F(x, y) = \sum_{n=0}^{\infty} y^{w(n)} x^n$. Also he noted that the coefficient of $x^n y^k$ in the expansion of

$$P(x, y) = 1/(1 - yx)(1 - yx^2)(1 - yx^4) \cdots \tag{2}$$

is the number of ways that n is the sum of k powers of 2, repetitions allowed. Therefore

$$P(x, y) = 1 + \sum_{n=1}^{\infty} \sum_{k=1}^{n} b(n, k) x^n y^k.$$

Also we obtain $1/P(x, y) = F(x, -y) = \sum_{n=0}^{\infty} (-y)^{w(n)} x^n$. Hence, the indicated relation is verified.

When $y = -1$, we have $P(x) = 1/F(x)$, which gives a relation between par-

titions into an odd number of parts and partitions into an even number of parts.

Also solved by L. Carlitz, N. G. Fine, M. G. Greening (Australia), Douglas Lind, E. F. Schmeichel, and the proposer.

ADVANCED PROBLEMS

All solutions of Advanced Problems should be sent to J. Barlaz, Rutgers—The State University, New Brunswick, NJ 08903. Solutions of Advanced Problems in this issue should be typed (with double spacing) on separate, signed sheets and should be mailed before September 30, 1970. Contributors (in the United States) who desire acknowledgement of receipt of their solutions are asked to enclose self-addressed, stamped postcards.

An asterisk (*) means neither the proposer nor the editors supplied a solution.

5740. *Proposed by L. R. Etzweiler and W. C. Waterhouse, Cornell University*

Let R be a commutative ring, S a multiplicative subset of R. Let M and N be R-modules. There is a canonical map

$$S^{-1}\operatorname{Hom}_R(M, N) \to \operatorname{Hom}_{S^{-1}R}(S^{-1}M, S^{-1}N).$$

This map is known to be injective if M is finitely generated, and bijective if M is finitely presented (N. Bourbaki, *Algèbre Commutative*, Ch. II, Section 2, Prop. 19). Find an example in which M is finitely generated and the map is not surjective.

5741. *Proposed by Simeon Reich, Israel Institute of Technology*

Let $p(z)=a_0+a_1z+\cdots+a_nz^n$, $a_n\neq 0$, be a polynomial with complex coefficients. It is known (Morris Marden, *The Geometry of the Zeros*, AMS, New York, 1949, p. 98) that if z_1 is the zero of largest modulus, then

$$|z_1| < \left\{1+\sum_{j=0}^{n-1}|a_j/a_n|^2\right\}^{1/2}.$$

Prove that if z_n is the zero of smallest modulus, then

$$|z_n| < \left\{1+\frac{1}{n}\left(\sum_{j=0}^{n-1}|a_j/a_n|^2\right)\right\}^{1/2}.$$

5742. *Proposed by Anne Penfold Street, University of Alberta*

Let G be a group and A a subgroup of G. Let $x\in G$, $x\notin A$. We say x *augments* A if $A_x\equiv A\cup(x, x^{-1})$ is also a subgroup of G. Suppose A is a subgroup of G such that any element of G augments A. Characterize G.

5743. *Proposed by Bjarni Jonssen and J. B. Nation, Vanderbilt University*

Let $f(x)$ be a real valued function with at least k derivatives. Given that for some real number r,

$$\lim_{x\to\infty} x^r f(x) = 0 \quad\text{and}\quad \lim_{x\to\infty} x^r f^{(k)}(x) = 0,$$

prove that $\lim_{x\to\infty} x^r f^{(j)}(x)=0$, $0\leqq j\leqq k$.

5744. *Proposed by Jan Mycielski, University of Colorado*

A cube $20\times 20\times 20$, is built out of bricks of the form $2\times 2\times 1$. The faces of the bricks are parallel to the faces of the cube but they need not all lie flat. Prove that the cube can be pierced by a straight line perpendicular to one of the faces which does not pierce any of the bricks.

5745.* *Proposed by Daniel Pedoe, University of Minnesota*

Q is a non-specialized quadric hypersurface in Euclidean space of n dimensions, L a hyperplane which intersects Q in a non-specialized quadric of one less dimension. We consider the pencil of quadric hypersurfaces $Q+kL^2=0$, and suppose that Q_1, Q_2, Q_3, Q_4 are any four of them; T_1, T_2, T_3, T_4 the total Gaussian curvatures at any point on the quadric of contact $Q=L=0$; M_1, M_2, M_3, M_4 the mean Gaussian curvature of the respective quadrics at the same point. Prove that the two cross-ratios

$$\{T_1, T_2, T_3, T_4\},\quad \{M_1, M_2, M_3, M_4\}$$

are equal. Deduce that if two conics touch at the two points P_1, P_2, then the ratio of their curvatures is the same at both points.

SOLUTIONS OF ADVANCED PROBLEMS

Homomorphism in a Semigroup

5522 [1967, 1014; 1968, 801]. *Proposed by T. C. Brown, Kiev State University, U. S. S. R.*

Let S be a periodic semigroup whose idempotents lie in the center. Then the mapping $x\to x^{n(x)}$, where $x^{n(x)}$ is idempotent, is a homomorphism.

II. *Solution by the proposer.* Let $x, y\in S$; choose m so that x^m, y^m, $(xy)^m$ are all idempotent. Then

$$(xy)^m = (xy)^m(xy)^m = x^2\omega(xy)^m.$$

By induction, $(xy)^m=x^m\omega'(xy)^m$. Similarly $(xy)^m=(xy)^m\omega''y^m$. Hence $(xy)^m=(xy)^m x^m y^m$. Now

$$(xy)^m = (xy)^m y^{m-1}yx^m x^m y^m = (xy)^m(y^{m-1}x^{m-1})(xy)x^m y^m.$$

By induction,

$$(xy)^m = (xy)^m(y^{m-1}x^{m-1})^m(xy)^m x^m y^m = (xy)^m(y^{m-1}x^{m-1})^m x^m y^m.$$

Then

$$\begin{aligned}(xy)^m &= (xy)^{m-1}(xy)(y^{m-1}x^{m-1})(y^{m-1}x^{m-1})^{m-1}x^m y^m\\ &= (xy)^{m-1}(y^{m-1}x^{m-1})^{m-1}x^m y^m.\end{aligned}$$

By induction, $(xy)^m=x^m y^m$.

EDITORIAL NOTE. The first solution printed for this problem [1968, 801] assumed a cancellation law. The proposer's solution above avoids the assumption.

Evaluation of $\int \prod_j \frac{\sin k_j(x-a_j)}{x-a_j}\,dx$.

5529 [1967, 1015; 1968, 914]. *Proposed by D. S. Mitrinović, University of Belgrade, Yugoslavia*

Evaluate

$$\text{Evaluation of} \int_{-\infty}^{\infty} \prod_{j=1}^{n} \frac{\sin k_j(x-a_j)}{x-a_j}\,dx,$$

with k_j, a_j, $j=1, 2, \cdots, n$ real numbers.

Note. The published solution for this problem is in error. Murray S. Klamkin remarks that it is to be expected that the given integral depend on all the k's and be symmetric in k_j, a_j. The formula obtained in the solution

$$I = \pi \prod_{j=2}^{n} \frac{\sin k_j(a_{j-1}-a_j)}{a_{j-1}-a_j}$$

does not involve k_1 and is not symmetric as required. ($k_1=0$ must imply $I=0$.)

Accordingly the solution is withdrawn and we urge our readers to reconsider the problem.

A Closed Subset in the Gelfand Space of a Banach Algebra

5654 [1969, 200]. *Proposed by L. J. Lardy and J. A. Lindberg, Jr., Syracuse University*

Let A be a commutative Banach algebra with identity, and let Φ_A denote the Gelfand space of A. Call a closed subset I of Φ_A an inverting set for A if the invertible elements of A are precisely those elements whose Gelfand transform does not vanish on I. Is the intersection of two inverting sets an inverting set?

Solution by Larry Eifler, Louisiana State University. Not necessarily. Let Δ denote the closed unit disk in $\mathbb{C}$. Let A be the uniform closure in $C(\Delta\times\Delta)$ of the polynomials in two variables. Then the Gelfand space of A can be identified with $\Delta\times\Delta$. Consider $K=\{(z, w)\in\Delta\times\Delta: |z|=|w|\}$ and $L=\{(z, w)\in\Delta\times\Delta: |z|=1 \text{ or } |w|=1\}$. Both K and L are inverting sets for A but $K\cap L$ is not an inverting set for A.

Also solved by the proposer.

Tangents to Rectifiable Curves

5659 [1969, 309]. *Proposed by K. L. Singh, Memorial University of Newfoundland*

Prove that a continuous rectifiable curve in a uniformly convex space possesses a tangent at almost all of its points.

Solution by H. Pëtard, III, Princeton University. We prove the more general result: A continuous rectifiable curve in a Banach space E with separable dual space E' has a tangent vector at almost all of its points.

Let $\gamma(t)$ be a continuous rectifiable curve in a Banach space E, with dual E' and second dual E''. Define a measure μ on $[0, 1]$ by setting $\mu([a, b])=$ variation of $\gamma(t)$ in $[a, b]$; μ is finite and positive. We are going to construct a function f, from $[0, 1]$ into E'', such that $\gamma(b)-\gamma(a)=\int_a^b f(t)\mu(dt)$ for $0<a<b<1$.

For each vector $x'\in E'$, set $\lambda_{x'}(A)=\int_A x'(d\gamma(t))$, where A denotes an arbitrary Borel subset of $[0, 1]$; $\lambda_{x'}$ is a measure satisfying the inequality $|\lambda_{x'}(A)| \leqq \|x'\|\cdot\mu(A)$ for any A. Hence, the Radon-Nikodým theorem tells us that $\lambda_{x'}(A)=\int_A f_{x'}(t)\mu(dt)$, where $|f_{x'}(t)| \leqq \|x'\|$ almost everywhere. If x', $y'\in E'$, then clearly $f_{x'+y'}(t)=f_{x'}(t)+f_{y'}(t)$ for almost all t. Therefore, every t outside of an appropriate union of exceptional sets gives rise to an element $f(t)\in E''$ satisfying $x'(f(t))=f_{x'}(t)$. If E' is separable, then our bad set of t's will be a countable union of exceptional sets, and therefore has measure zero. Clearly then, $\|f(t)\|\leqq 1$ almost everywhere, and

$$x'(\gamma(b) - \gamma(a)) = \int_a^b x'(f(t))\mu(dt),$$

which implies the desired identity $\gamma(b)-\gamma(a)=\int_a^b f(t)\mu(dt)$.

Now we can write $\mu(A)=\mu_s(A)+\int_A g(t)dt$, where $g\in L^1(0, 1)$ and μ_s is singular with respect to Lebesgue measure. Therefore,

$$\begin{aligned}\gamma(b) - \gamma(a) &= \int_a^b f(t)g(t)dt + \int_a^b f(t)\mu_s(dt)\\ &\equiv \int_a^b f(t)g(t)dt + \lambda_s((a, b]).\end{aligned}$$

Here λ_s is a vector-valued measure satisfying $\|\lambda_s(A)\| \leqq \mu_s(A)$.) So

$$(*)\qquad \begin{aligned}\lim_{\epsilon\to 0} \epsilon^{-1}(\gamma(a + \epsilon) - \gamma(a)) &= \lim_{\epsilon\to 0} \epsilon^{-1}\int_a^{a+\epsilon} f(t)g(t)dt\\ &\quad + \lim_{\epsilon\to 0} \epsilon^{-1}\lambda_s((a, a + \epsilon]).\end{aligned}$$

The first limit on the right-hand side of (*) exists for almost every $t\in[0, 1]$, since Lebesgue's theorem on differentiation of the integral holds for vector-valued functions (see Dunford-Schwartz p. 217 for the proof, which just copies the proof for the scalar case). The final term in (*) is zero for almost all t, since $\|\epsilon^{-1}\lambda_s((a, a+\epsilon])\| \leqq \epsilon^{-1}\mu_s((a, a+\epsilon])\to 0$ as $\epsilon\to 0$.

Thus, equation (*) shows that γ has a tangent vector for almost every t, provided only that the dual space E' is separable.

Finally, suppose that γ takes its values in any reflexive Banach space E. To show that γ has tangents, we may restrict our attention to the subspace

$E_0 \subseteq E$, the closed linear span of the vectors $\{\gamma(t): 0 \leqq t \leqq 1\}$. E_0 is a separable reflexive Banach space, which implies that E_0' is separable.

A Combinatorial Iden'ity

5663 [1969, 309]. *Proposed by D. Ž. Djoković, University of Waterloo, Ontario*

Show that

$$\sum \frac{1}{p_1!p_2! \cdots p_n!} = \frac{1}{k!}\binom{n-1}{k-1},$$

where the summation is over all nonnegative integers $p_1, p_2, \cdots$, such that $p_1+2p_2+ \cdots +np_n=n$, $p_1+p_2+ \cdots +p_n=k$.

Solution by Henry Ricardo, Yeshiva University and Manhattan College. This result is part of the more general problem E 700 [1946, 340]. A solution, essentially that given for E 700, follows: In the formal identity

$$\begin{aligned}(x + x^2 + \cdots + x^n)^k &= x^k(1 - x^n)^k(1 - x)^{-k} \\ &= x^k(1 - x^n)^k \sum_{j=0}^{\infty}\binom{k+j-1}{j}x^j,\end{aligned}$$

we equate the coefficients of x^n to get

$$\sum \frac{k!}{p_1!p_2! \cdots p_n!} = \binom{n-1}{k-1},$$

which is the desired result upon dividing by $k!$. Here we have used the multinomial theorem where the summation and the p_i have the same meaning as in our problem.

Also solved by H. D. Abramson (Scotland), Anders Bager (Denmark), M. T. L. Bizley (England), Michel Bousquet, Robert Breusch, Paul Brock, L. Carlitz, C. A. Church, Jr., S. C. Currier, Jr., A. E. Fekete, N. J. Fine, Roberto Frucht (Chile), H. W. Gould, M. G. Greening (Australia), D. A. Hejhal, M. Hirsello, J. E. Kilpatrick, M. S. Klamkin, Harry Lass, C. P. Lawes, Douglas Lind, O. P. Lossers (Netherlands), Andrzej Mąkowski (Poland), Marvin Marcus, S. E. Payne, Henri Pétard, Jernej Polajnar (Yugoslavia), D. P. Roselle, Barry Simon, James Singer, T. Tamura (Japan), E. W. Trost (Switzerland), D. A. Zave, David Zeitlin, and P. J. Zwier.

Note. This problem has appeared earlier. For other solutions and earlier references see *Siam Review*, Vol. 9 (1967) p. 252.

The Inverse Cube Law for a Planet

5664 [1969, 310]. *Proposed by Harry Pollard, Purdue University*

Suppose a particle is attracted to a fixed center O by a force proportional to the inverse cube of its distance r from O. Without determining the possible orbits, show that one of these three things must occur:

(a) the particle moves in a circle;

(b) the particle collides with O in a finite time;

(c) $r \to \infty$ as $t \to \infty$.

Conclude that the inverse cube law is a poor substitute for the inverse square law in designing a solar system.

I. *Solution by M. L. Laplaza, University of Puerto Rico.* It is well known that we have motion in a plane when the particle is subject to a central force. The hypothesis becomes

$$\frac{d^2x}{dt^2} = k\frac{x}{(x^2+y^2)^2}, \qquad \frac{d^2y}{dt^2} = k\frac{y}{(x^2+y^2)^2}.$$

Suppose that $x^2+y^2 \neq 0$; then

$$\frac{d}{dt}\left[\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2\right] = \frac{d}{dt}\left(-k\frac{1}{x^2+y^2}\right)$$

and so,

$$\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 = k_1 - \frac{k}{x^2+y^2}, \qquad \frac{d}{dt}\left(x\frac{dx}{dt} + y\frac{dy}{dt}\right) = k_1.$$

From this, there follow

$$x\frac{dx}{dt} + y\frac{dy}{dt} = k_1t + k_2, \qquad \frac{d}{dt}(x^2+y^2) = 2k_1t + 2k_2,$$

$$x^2 + y^2 = k_1t^2 + 2k_2t + k_3.$$

If $\lim_{t\to\infty}(x^2+y^2) \neq \infty$, then $k_1 = k_2 = 0$, and $x^2+y^2 = k_3$. If k_1 or $k_2 \neq 0$, x^2+y^2 may become zero for some value of t, but eventually goes to infinity.

II. *Solution by Robert Israel, University of Chicago.* It is well known that for central-force problems the particle motion is in a plane, that the angular momentum $A = mv_0r$ is constant, and that

$$\frac{d^2r}{dt^2} = \frac{v_0^2}{r} + \frac{F}{m}.$$

Substitution of $F = -k/r^3$ leads to the equation

$$\frac{d^2r}{dt^2} = \frac{A^2 - km}{m^2r^3}.$$

Thus if $A^2 = km$, the radial component of velocity is constant, so the particle either reaches the origin in a finite time, stays a constant distance from the center (travelling in a circular orbit since A is constant and nonzero), or recedes from the center at a constant rate. If $A^2 > km$, then $r \to \infty$ as $t \to \infty$ because conservation of energy prevents the particle from colliding with the origin and because the assumption that r is bounded by some R gives a lower limit of $(A^2 - km)/m^2R^3$ for d^2r/dt^2, which implies that the particle must attain $r > R$

in a finite time. If $A^2 < km$, the particle may encounter the origin in a finite time or go infinitely far away as $t \to \infty$, because if the motion is bounded by some R we have a lower limit of $(km - A^2)/m^2R^3$ for $(-d^2r/dt^2)$, so that in a finite time the radial velocity must be negative, after which in a finite time the particle must encounter the origin.

The inverse-cube law is a poor choice for a solar system because the circular orbits are not stable; almost any slight perturbation will send a planet on a path leading either into the sun or out of the system.

Also solved by L. E. Clarke (England), D. H. Gaeddert, D. A. Hejhal, O. P. Lossers (Netherlands), Henrik Meyer (Denmark), H. Pétard, Jr., and J. S. Shipman.

Clarke furnishes the reference to L. A. Pars, *Introduction to Dynamics* (Cambridge, 1953), p. 254 for the result: A circular orbit described under an inverse nth power law of attraction is stable if and only if $n < 3$.

Fixed Points in a Limited Contraction

5672 [1969, 565]. *Proposed by Stephen Weingram, Purdue University*

A continuous map $f: R^n \to R^n$ is a contraction outside the set K if there is a constant $c(0 < c < 1)$ such that for any two points x, y outside K, $|f(x) - f(y)| \leqq c|x - y|$. Prove that if $f: R^n \to R^n$ is a contraction outside a compact set K, it has a fixed point.

Solution by Michael Menn, Boston College. Assume (without loss of generality) that $0 \notin K$. Let $M = \max\{\text{lub}|f|/K, |f(0)|/(1-c)\}$ and let B be the closed ball of radius M about the origin. Note that $cM + |f(0)| \leqq M$. The result will follow from the Brouwer fixed point theorem if it can be shown that f maps B into B. But if $x \in K$ then $f(x) \in B$, and if $x \in B - K$ then $|f(x)| \leqq |f(x) - f(0)| + |f(0)| \leqq c|x| + |f(0)| \leqq cM + |f(0)| \leqq M$.

Also solved by D. T. Adams, D. R. Anderson & V. M. Sehgal, Einar Andresen (Norway), David Boyd, Theodore Chang, E. P. Del Norte, D. A. Herrero, R. J. Loy, Joel Spencer & Emmett Keeler, Jean Spitzer, R. A. Struble, J. H. van Lint (Netherlands), D. A. Zave, Martin Ziegler (Germany), and the proposer.

Boyd proves the result using the condition on f that $|f(x)| \leqq A + c|x|$, $A \geqq 0$, $0 < c < 1$ outside a bounded set K. Loy notes the necessity of the boundedness condition on K and that K may or may not contain any of the fixed points. However, if $f(K) \subseteq K$, then all of the fixed points are in K.

Product of Algebraic Numbers

5674 [1969, 565]. *Proposed by L. Carlitz, Duke University.*

It is proved in problem 5542 [1968, 1021] that the following statement is incorrect: If a, b, are algebraic over F of degree m and n respectively and if m and n are relatively prime, then ab is algebraic over F of degree mn.

Show that the statement is correct when $F = \mathrm{GF}(q)$, $q = p^n$, p prime, $n \geqq 1$.

Solution by H. F. Mattson, Jr., Sylvania Applied Research Laboratory, Waltham, Mass. We use the fact that greatest common divisors satisfy

$$(q^x - 1, q^y - 1) = q^{(x,y)} - 1$$

which one proves easily by looking at the fields of the indicated sizes. To say that a has degree m means that a is in $GF(q^m)$ but not in any proper subfield. This means that a is a root of unity of order α, where α divides q^m-1; but α does not divide q^d-1 for any proper divisor d of m. Similarly for b of order β and degree n.

Now ab is a root of 1 of order l.c.m.(α, β). It has degree D for some D dividing mn. Thus, both α and β divide q^D-1. Thus, α divides $(q^D-1, q^m-1) = q^{(D,m)}-1$, which implies m divides D. Similarly, n divides D. Since $(m, n)=1$, $D=mn$ as required.

Also solved by Anders Bager (Denmark), Douglas Lind, R. A. Moore, J. H. Smith, T. Tamura (Japan), and the proposer.

Bager notes that the two n's appearing in the problem are not to be taken as equal.

REVIEWS

EDITED BY KENNETH O. MAY

COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER

Printed materials for review should be sent to Kenneth O. May, American Mathematical Monthly, Department of Mathematics, University of Toronto, Toronto 5, Canada. Films and correspondence relating to films should be sent to Seymour Schuster, Carleton College, Northfield, Minnesota 55057.

All unsigned material is written by the editors. A boldface capital **C** *in the margin indicates that a review is based in part on classroom use. Professors willing to write such a review should first inform the editor in order to avoid duplication.*

C *Basic Statistics.* By David Blackwell. McGraw-Hill, New York, 1969. v+143 pp. \$5.50. (Telegraphic Review, October 1969.)

This book is truly a new approach for a course in statistics to satisfy a general requirement in mathematics for undergraduates. It is one text in which the author accomplishes what he says he will in the preface.

Used in the classroom during the summer of 1969, it was delightful for both instructor and students. One is first attracted by the cover, the small size of the book, the largeness and clearness of the print, and the eye-catching diagrams. One is next impressed with the substance and depth of the subject matter, covering the usual topics, but covering them in a logical, well-organized, unique and intriguing manner. The material of one chapter is subtly structured to lead into that of the next chapter. The instructor finds himself unable to change the order of presentation, but not wanting to make a change. Problems at the end of each chapter also contribute to the smooth movement to the next chapter and include ones that challenge the more capable student to investigate the

THE AMERICAN MATHEMATICAL MONTHLY

(FOUNDED IN 1894 BY BENJAMIN F. FINKEL)

THE OFFICIAL JOURNAL OF

THE MATHEMATICAL ASSOCIATION OF AMERICA, INC.

VOLUME 77 NUMBER 7

CONTENTS

(continued on inside cover)

AUGUST-SEPTEMBER 1970

NOTICE TO AUTHORS

Articles should be typewritten on good quality paper, triple spaced with wide margins. Submit the original and a duplicate; keep a complete copy for protection against loss. Please follow the format used in current issues of the MONTHLY. The manuscript must make the author's intent clear to the printer. A paper accepted for publication by the editor but not acceptable to the printer will be returned to the author for retyping. Please use the American Mathematical Society *Manual for Authors of Mathematical Papers* as your guide and please conform to correct English usage. A *Manual for Monthly Authors* is available on request.

Prospective MONTHLY authors are advised to consult the Statement of Policy in the January 1969 issue, p. 2. Specialized research is usually unsuitable. Backlog: Main Articles 11 months, Math. Notes 10 months, Research Problems 7 months, Classroom Notes 7 months, Math. Education 8 months.

EDITORIAL CORRESPONDENCE AND MAIN ARTICLES: to HARLEY FLANDERS, American Mathematical Monthly, Tel Aviv University, Ramat Aviv, Israel (see Notice, p. 555); NOTES, etc.: to the corresponding Associate Editor; ADVERTISING CORRESPONDENCE: to RAOUL HAILPERN, Mathematical Association of America, SUNY at Buffalo, Buffalo, N. Y. 14214; CHANGE OF ADDRESS and SUBSCRIPTIONS: to A. B. WILLCOX, Mathematical Association of America, 1225 Connecticut Ave., N.W., Washington, D. C. 20036.

Annual dues for members of the Association (including a subscription to the American Mathematical Monthly) are $10.00. For nonmembers the subscription price is $15.00.

PUBLISHED BY THE ASSOCIATION at Washington, D. C., and Menasha, Wisconsin during the months of January, February, March, April, May, June–July, August–September, October, November, December.

Second-class postage paid at Washington, D. C., and additional mailing offices. Acceptance for mailing at special rate of postage provided for in the Act of February 28, 1925, embodied in Paragraph 4, Section 538, P. L. and R., authorized April 1, 1926.

PRINTED IN THE UNITED STATES OF AMERICA

WHAT IS A SHEAF?

J. ARTHUR SEEBACH, JR., LINDA A. SEEBACH, AND LYNN A. STEEN,
St. Olaf College

Ever since Jean Leray and Henri Cartan in 1950 formally introduced the concept of a sheaf, the various examples and applications of sheaves have come to play a major role in such diverse fields as several complex variables, algebraic geometry, and differential and algebraic topology. Yet nearly all monographs which use or introduce sheaves assume the sophistication of graduate level algebraic topology. So it is very difficult for an undergraduate to acquire from the available literature a real understanding of sheaves and their applications. It is the purpose of this article to introduce the theory of sheaves at an elementary level with the hope that the interested reader will then be able to approach any of the standard treatises (e.g., [2], [3], [6], [10], or [11]) with significant insight.

Our avenue of approach to the theory of sheaves will be through examples drawn from three major areas of mathematics: from analysis, the sheaf of germs of holomorphic functions; from algebra, the sheaf of local rings; and from geometry, the sheaf of differential forms. We will develop each of these particular sheaves in considerable detail, for the different perspectives thus revealed will more readily make transparent the subsequent discussion of the general theory of sheaves.

1. The sheaf of germs of holomorphic functions. A *holomorphic* (or *analytic*) function on an open subset D of complex n-space C^n is defined to be a complex valued function on D which has a local power series representation at each point of D. Osgood's lemma [6, p. 2] asserts that a continuous function on $D \subset C^n$ is holomorphic if and only if it is holomorphic in each variable separately.

A very important property of holomorphic functions is that they are uniquely determined by their behavior on open sets: if f and g are holomorphic on a domain D (a *domain* is a connected open set), and if f equals g on a nonempty open subset of D, then f equals g on all of D. To see this, we need only observe that the largest open subset of D on which $f=g$ is also closed (relative to D), since the partial derivatives which determine the power series expansion are continuous. Since D is connected, this set must be D.

Now if $z \in C^n$, we say that f is *holomorphic at z* if it is holomorphic on some neighborhood of z. The collection A_z of functions holomorphic at z forms an algebra over the field of complex numbers in which the operations of sum and

Professor Steen received his Ph.D. in 1965 under Kenneth Hoffman at M.I.T.; Professor Seebach received his Ph.D. in 1968 under A. I. Weinzweig at Northwestern; his wife Linda Seebach is a Ph.D. candidate under Daniel Zelinsky, also at Northwestern. Professors Steen and Seebach are the authors of the recently published *Counterexamples in Topology.*

The paper printed here was derived from a series of lectures presented by the authors at a weekly joint colloquium of the mathematics staffs of St. Olaf and Carleton Colleges, both located in Northfield, Minnesota. *Editor.*

product involve intersection of domain: if $f: U \to C$ and $g: V \to C$, then $f+g: U \cap V \to C$ and $fg: U \cap V \to C$. We let I_z be the ideal in A_z consisting of those functions in A_z which vanish identically on some neighborhood of z.

The *algebra of germs of holomorphic functions* at z is then defined to be the quotient ring (algebra) A_z/I_z, and is denoted by O_z. So a germ of a holomorphic function is an element $f+I_z$ of O_z, where f is holomorphic at z. We will usually denote this germ by $[f]_z$. Following the usual practice, we shall often identify, or fail to distinguish between, two functions which belong to the same germ. This sloppiness is somewhat justified by the uniqueness property stated above, for two functions which belong to the same germ and are defined on the same domain D must differ by a function in I_z, which means that they agree on some neighborhood of z, and thus must agree on D.

We may now define the *stalk space* (*espace étalé*) of germs of holomorphic functions to be the set $S = \{(z, [f]_z) \mid f \text{ is holomorphic at } z \in C^n\}$ together with the natural mapping ρ from S to C^n defined by $\rho((z, [f]_z)) = z$. We call $\rho^{-1}(z)$ the *stalk* at the point $z \in C^n$; it is simply a copy of O_z, the algebra of germs of holomorphic functions at z. The stalk space S is thus the disjoint union of the stalks. Intuitively, we shall picture S as a space of interpenetrating sheets lying over C^n, with ρ projecting S onto C^n (Fig. 1).

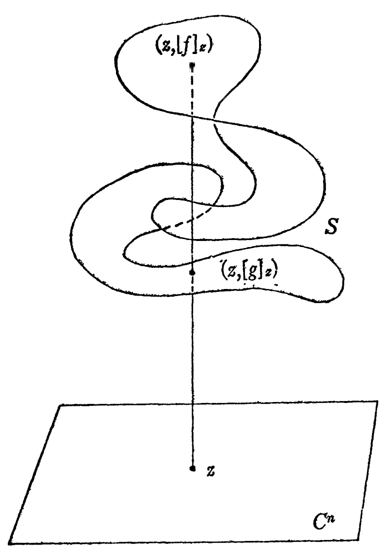


FIG. 1

To make this intuitive picture more precise, we lift the topology of C^n back to S, to make S into a topological space. For each open set U in C^n and each function f which is holomorphic on U, we define $V(f, U) = \{(z, [f]_z) \mid z \in U\}$. Each such $V(f, U)$ is contained in S, and the collection of all such sets covers S, for if $(z, [f_0]_z) \in S$, f_0 must be holomorphic on some neighborhood U_0 of z and $(z, [f_0]_z) \in V(f_0, U_0)$. Furthermore, $V(f_1, U_1) \cap V(f_2, U_2) = V(f, U)$ where $U = \{z \in U_1$

$\cap U_2 | [f_1]_z = [f_2]_z\}$ and $f = f_1|_U = f_2|_U$. Thus the sets $V(f, U)$ form a basis for a topology on S, and relative to this topology, the projection ρ is a *local homeomorphism*. That is, for each basis neighborhood $V(f, U)$ in S, the one-to-one map $\rho|_{V(f,U)}$ is a homeomorphism onto U. For if we let $\rho_{f,U}$ denote $\rho|_{V(f,U)}$, and if N is an open subset of U, then $\rho_{f,U}^{-1}(N) = V(f, N)$ which is open in S, while if $V(f, U') \subset V(f, U)$, then $\rho_{f,U}(V(f, U')) = U'$. The topology on S is uniquely determined by the requirement that the projection ρ be a local homeomorphism.

This topological space S, together with the local homeomorphism ρ which projects S onto C^n, is called the *sheaf of germs of holomorphic functions* over the *base space* C^n. As the agricultural terminology implies, we think of the sheaf as a bundle of stalks (Fig. 2), each with a full head of germs (or, if you wish, seeds, or grain).

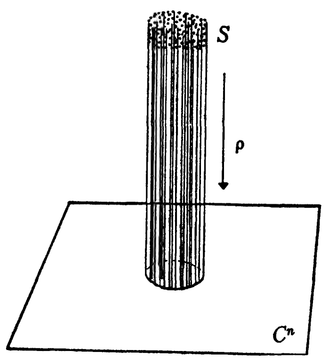


FIG. 2

We can show that the stalk space S is Hausdorff as follows: points of S may differ either because they are on different stalks, or because they are on different levels of the same stalk. In the first case, the projections of two points $p, q \in S$ differ in C^n; so, since C^n is Hausdorff, there are disjoint neighborhoods of $\rho(p)$ and $\rho(q)$ which may be lifted back to S. To be specific, if $p = (z, [f]_z)$ and $q = (w, [g]_w)$ where $z \neq w$, then there exist disjoint open neighborhoods U_z, U_w of z and w respectively and on them holomorphic functions $f \in [f]_z$ and $g \in [g]_w$, respectively, so that $V(f, U_z)$ and $V(g, U_w)$ are disjoint neighborhoods of p and q.

The second case is a bit more complex, since it depends on the uniqueness property of holomorphic functions. If $p = (z, [f]_z)$ and $q = (z, [g]_z)$ are different points on the same stalk, then $[f]_z \neq [g]_z$; so there must exist different holomorphic functions $f \in [f]_z$ and $g \in [g]_z$ which are both defined on some neighborhood U of z. We claim that $V(f, U)$ and $V(g, U)$ are then disjoint neighborhoods of p and q, for if $(w, [h]_w) \in V(f, U) \cap V(g, U)$, then $w \in U$ and $[f]_w = [h]_w = [g]_w$. But as we observed above, the uniqueness property of holomorphic functions implies that the two functions f and g with the same domain U which belong to the same germ $[h]_w$ must be identical on U. But they are not identical on U, since $[f]_z \neq [g]_z$. So $V(f, U) \cap V(g, U) = \varnothing$, and thus S is Hausdorff.

There is still another consequence of the uniqueness property of holomorphic functions that can be used to further illuminate the sheaf of germs of holomorphic functions. The uniqueness property may be roughly interpreted as saying that the global behavior of a holomorphic function is uniquely determined by its behavior on any open set. This makes meaningful the vague question of identifying the largest domain to which a given holomorphic function can be extended. In the classical study of analytic functions this question led to the concept of a Riemann surface, or more generally to complex analytic manifolds.

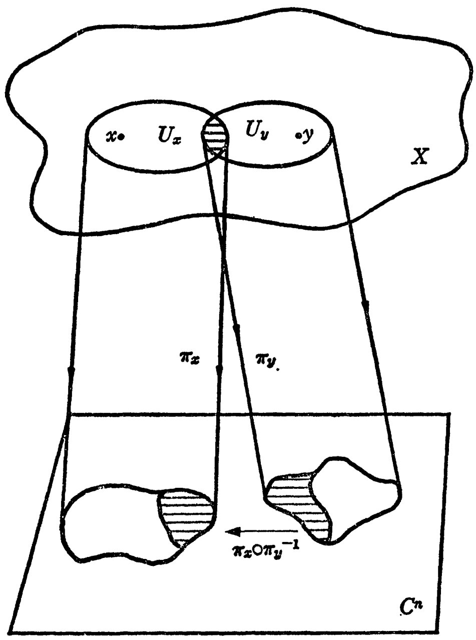


FIG. 3

A manifold is, essentially, a topological space which is locally homeomorphic to complex Euclidean n-space C^n. To be more precise, we will call a topological space X *locally Euclidean* (of dimension n) if every $x \in X$ is contained in an open set U_x which is homeomorphic under a mapping π_x to some subset of C^n, where, furthermore, the *coordinate patches* U_x are coherent in the sense that for each $x, y \in X$, $\pi_x \circ \pi_y^{-1}$ is a homeomorphism between $\pi_y(U_x \cap U_y)$ and $\pi_x(U_x \cap U_y)$. The first half of this definition guarantees that X is locally like C^n, while the second condition requires that the locally Euclidean patches overlap so as to form a coherent Euclidean structure on all of X. We shall call each pair (U_x, π_x) a *local coordinate system*, since π_x^{-1} lifts the coordinate system of C^n back to U (Fig. 3).

Since the same topological space may be covered by several different collections of coordinate systems (U_x, π_x), and since we do not wish to distinguish

between two covers which provide essentially the same coordinate structure on X, we define a *manifold* to be a locally Euclidean topological space in which the collection of coordinate systems (U_x, π_x) is maximal with respect to the defining properties for a locally Euclidean space. Since each locally Euclidean space generates a unique manifold, we shall often refer to locally Euclidean spaces as manifolds even if the collection of local coordinate systems is not maximal. Other types of manifolds may be produced by projecting to real Euclidean space R^n instead of to C^n or by requiring that the homeomorphisms $\pi_x \circ \pi_y^{-1}$ be analytic or C^∞ (infinitely differentiable); such manifolds are naturally called *analytic manifolds* or *C^∞ manifolds.*

A function $f: X \to Y$ from one analytic manifold to another is called *holomorphic* if for each x and y, $\pi_y \circ f \circ \pi_x^{-1}$ is holomorphic on its domain, which is $\pi_x(U_x \cap f^{-1}(U_y))$. In the special case, where $Y = C^1$, the identity map $i: Y \to C^1$ is used to define the local coordinate systems. So a holomorphic function f from the analytic manifold X to C^1 is characterized by the property that $f \circ \pi_x^{-1}$ is holomorphic on $\pi_x(U_x)$.

It should be clear from this description that the sheaf of germs of holomorphic functions can be regarded as an analytic manifold, using the projection ρ to define the local coordinate systems. It is a particularly important manifold, since on it we can define what is known as the *universal holomorphic function.* This is the mapping $F: S \to C$ defined by $F((z, [f]_z)) = f(z)$. F is clearly holomorphic since for each local coordinate system $(U, \rho|_U)$, we have $F \circ (\rho|_U)^{-1} = f|_{\rho(U)}$ where $f: \rho(U) \to C$ is holomorphic.

$$\begin{tikzcd} S \arrow[dr, "F"] & \\ D \arrow[u, "\hat{f}"] \arrow[r, "f"'] & C \end{tikzcd}$$

FIG. 4

F is universal in the sense that the behavior of all holomorphic functions on C^n is subsumed in that of F. In particular, whenever $f: D \to C$ is holomorphic (where D is a domain in C^n), we can factor f through the sheaf S as follows: there exists a unique function $\hat{f}: D \to S$ such that $F \circ \hat{f} = f$. Clearly $\hat{f}$ is defined by $\hat{f}(z) = (z, [f]_z)$, so the associated diagram (Fig. 4) is commutative; $\hat{f}$ is continuous since $\hat{f}^{-1}(V(f, U)) = U$.

With this structure, we can now describe the *domain of holomorphy* of a given holomorphic function f—that is, the largest domain to which f can be

uniquely extended. It is the connected component E of S which contains $\hat{f}(D)$. (Of course $\hat{f}(D)$ is connected since it is the continuous image of a connected set.) Although D is not literally a subset of E, it is imbedded by $\hat{f}$ in E, and thus the universal holomorphic function F is the extension to E of the function f.

2. The sheaf of local rings. Let A be a commutative ring with 1, and S a multiplicatively closed nonempty subset of A with $0 \notin S$. We construct from A and S a ring A_S called a *ring of quotients of* A, in which the elements of S have multiplicative inverses. On the set

$$A \times S = \{(a, s) \mid a \in A, s \in S\}$$

we define an equivalence relation $(a, s) \sim (b, t)$ if and only if there exists $r \in S$ such that $(at-bs)r=0$. We also define two operations,

$$(a, s) + (b, t) = (at + bs, st) \quad \text{and} \quad (a, s)(b, t) = (ab, st),$$

which are compatible with the relation. We denote by A_S the ring of equivalence classes with the induced operations. As in the ring of integers with the set of nonzero elements as S, the equivalence class of (a, s) is denoted by a/s; thus we call S the *set of denominators*.

The 0 of A_S is $0/s$ (any s in S), the identity is s/s; and if $s \in S$, $s^{-1}=1/s$. There is a homomorphism $\alpha: A \to A_S$ defined by $\alpha(a)=as/s$, which is independent of the choice of s. Of course if A is an integral domain, α is one-to-one because Ker $\alpha = \{a \mid sa=0 \text{ for some } s \in S\}$.

If I is an ideal of A, the ideal $\alpha(I)$ in A_S can be represented by

$$\alpha(I) = \{a/s \mid a \in I, s \in S\},$$

and we shall write IA_S for $\alpha(I)$. This function α on the set of ideals of A defines a one-to-one correspondence between the set of prime ideals in A_S and the set of prime ideals in A whose intersection with S is empty [**12**, p. 223].

Since we may describe a prime ideal P in A as one whose complement is multiplicatively closed, we may form the ring of quotients of A whose set of denominators is the complement of P. We shall denote this ring of quotients by A_P, and call it the *local ring* of A at P. The ring A_P has only one maximal ideal, PA_P, since clearly P is the largest prime ideal of A with the property that its intersection with the complement of P is empty.

These local rings will be the stalks for the sheaf of local rings and the set of prime ideals of A will form the base space. This space is called the *spectrum* of A, denoted by Spec A, and is topologized by taking as a basis for the topology all sets $V_x = \{P \in \text{Spec } A \mid x \notin P\}$ where $x \in A$. Then $V_1 = \text{Spec } A$, $V_0 = \varnothing$, and $V_x \cap V_y = V_{xy}$; thus $\{V_x\}_{x \in A}$ is a basis. Since $\bigcup_{x \in M} V_x = \{P \mid (x)_{x \in M} \not\subset P\}$ where $(x)_{x \in M}$ is the ideal generated by the subset M of A, any ideal I of A defines an open set $V_I = \{P \mid I \not\subset P\}$, and every open set U is of this form although I is not uniquely determined by U. A closed set, then, is a set of primes containing some fixed ideal, so that a point P in Spec A is closed if and only if it is a maximal

ideal. For most rings, therefore, the base space is not even T_1; however, Spec A is always T_0.

It is possible to define a ring of quotients with respect to the complement of a prime P because it is multiplicatively closed, but the complement of a union of primes is also multiplicatively closed. Hence, with each nonempty open set U in Spec A we may associate the ring of quotients $A_U = \{a/s \mid P \in U \Rightarrow s \notin P\}$ whose set of denominators is the complement of the union of all the primes in U. If U and V are open sets in Spec A such that $U \subset V$, we may define *restriction homomorphisms* $\rho_{U,V}: A_V \to A_U$ as follows: if $a/s \in A_V$, s is not an element of any prime ideal P in V, and so *a fortiori* not an element of any prime ideal P in U. Hence a/s is also an element of A_U. We define $\rho_{U,V}(a/s) = a/s$, but this map is not the identity, or even one-to-one, since the equivalence classes which are used to define the ring A_U are larger than those used to define A_V. The kernel of $\rho_{U,V}$ consists of those elements a/s such that a is a zero divisor with respect to an element in one of the prime ideals of V which is not in any element of U.

An important property of $\rho_{U,V}$ is the commutativity of the diagram in Fig. 5, where $\alpha_U: A \to A_U$ takes a to $a/1$. Now $\rho_{U,V}$ is uniquely determined by this property and from this it follows that $\rho_{U,U}$ is the identity map and that if $U \subset V \subset W$, then $\rho_{U,W} = \rho_{U,V} \circ \rho_{V,W}$. This system, consisting of Spec A, the rings A_U, and maps $\rho_{U,V}: A_V \to A_U$ when $U \subset V$, is called a *presheaf* over Spec A. Besides the maps $\rho_{U,V}$ corresponding to pairs of open sets for which $U \subset V$, we

$$\begin{tikzcd} & A_V \arrow[dd, "\rho_{U,V}"] \\ A \arrow[ur, "\alpha_V"] \arrow[dr, "\alpha_U"'] & \\ & A_U \end{tikzcd}$$

FIG. 5

$$\begin{tikzcd} & A_V \arrow[dd, "\rho_{P,V}"] \\ A \arrow[ur, "\alpha_V"] \arrow[dr, "\alpha_P"'] & \\ & A_P \end{tikzcd}$$

FIG. 6

can define maps $\rho_{P,V}: A_V \to A_P$ when $P \in V$. If V is open and $P \in V$, we define $\rho_{P,V}: A_V \to A_P$, by $\rho_{P,V}(a/s) = a/s$, which is possible since $a/s \in A_V$ implies $s \notin P$. The map $\rho_{P,V}$ is the unique map which makes the associated diagram commute (Fig. 6). As before, it follows from the uniqueness of $\rho_{P,V}$ that

$$\rho_{P,W} \circ \rho_{W,U} = \rho_{P,U} \qquad \text{if } P \in W \subset U.$$

If U is open in Spec A, and u is any element of A_U, we may treat u as a function from U to the stalk space $S = \bigcup \{A_P \mid P \in \text{Spec } A\}$ by defining $u(P) = \rho_{P,U}(u) \in A_P$ for $P \in U$. If $V \subset U$, we call $\rho_{V,U}(u)$ the *restriction* of u to V. If $u \in A_U$ and $v \in A_V$, and if there is an open set $W \subset U \cap V$ such that $\rho_{W,U}(u) = \rho_{W,V}(u)$, we say u and v *agree on the open set* W, for if $P \in W$, then

$$u(P) = \rho_{P,U}(u) = \rho_{P,W} \circ \rho_{W,U}(u) = \rho_{P,W} \circ \rho_{W,V}(v) = \rho_{P,V}(v) = v(P).$$

If $u \in A_U$, either the function u or its image $u(U)$ in S is called a *section* of the presheaf. The set of sections $u(U)$ covers S, for if $a/s \in A_P$ then s is a denominator in A_V, where $V = \{Q \mid s \notin Q\}$, and $a/s \in A_V$ is a section over V whose image at P is a/s.

The collection of all these sections $u(U)$ is a basis for a topology on S, since the intersection of two sections is a union of sections. For, suppose $u \in A_U$ and $v \in A_V$ are sections over the open sets U and V. If $u(U) \cap v(V) = \varnothing$ there is nothing to show. If $x \in u(U) \cap v(V)$, $x = \rho_{P,U}(v) = \rho_{P,V}(v) \in A_P$ for some $P \in U \cap V$. Then $u = a/s$, where s is not an element of any element of U, and $v = a'/s'$, where s' is not an element of any element of V. Since $a/s = a'/s'$ in A_P, there exists $t \notin P$ such that $t(as' - a's) = 0$. Let $W = \{Q \mid t \notin Q\}$. Then $a/s = a'/s'$ in $A_{W \cap U \cap V}$ and, since the diagram in Fig. 7 commutes, the section over $W \cap U \cap V$ defined by $a/s = a'/s'$ is a subset of $u(U)$ and of $v(V)$ and it is a neighborhood of x.

$$\begin{tikzcd}
A_V \arrow[dr] \arrow[drr] & & \\
& A_{W \cap U \cap V} \arrow[r] & A_P \\
A_U \arrow[ur] \arrow[urr] & &
\end{tikzcd}$$

FIG. 7

In this topology the projection ρ is a local homeomorphism, for if $x \in A_P$ there is a section v over V through x for some V, and the restriction of ρ to the section $v(V)$ is one-to-one and onto. If $U \subset V$ is open and $x \in U$, the restriction of v to U is also a section through x, and thus $u(U)$, where $u = \rho_{U,V}(v)$, is open in $u(V)$; so ρ is continuous. If N is an open subset of $v(V)$ it is a union of sections over open subsets of V and $\rho(N)$ is the union of these open subsets.

The topological space S, together with the projection $\rho: S \to \operatorname{Spec} A$, is called the *sheaf of local rings* over Spec A. If U is an open subset of Spec A, any continuous function $f: U \to S$ such that $\rho \circ f$ is the identity on U is called a *section* of the sheaf S; the set of all sections over the open set U is denoted by $\Gamma(U, S)$. The relation between the sections of the sheaf S and the sections of the presheaf (that is, the elements $u \in A_U$) is rather subtle, for even though each presheaf section may be thought of as a (continuous) function on U which is a local inverse for ρ, two anomalies may occur. It may be that two different elements $u = a/s$ and $v = b/t$ of the ring A_U yield the same function under the interpretation outlined above, or there may be functions in $\Gamma(U, S)$ which cannot be derived from any section $u \in A_U$. Thus the interpretation map from A_U to $\Gamma(U, S)$ need not be either one-to-one or onto, though if U is a basis set (that is, one of the form V_x for some $x \in A$) this map is an isomorphism [**5**, No. 4, p. 86]. For this reason, the open sets of the form V_x are often called *distinguished* open sets.

If the base space Spec A is not Hausdorff, certainly the sheaf S cannot be Hausdorff. But even if Spec A is T_2 it may well happen (although it is difficult

to visualize it on Hausdorff paper) that two distinct sections over U, u and u', may agree on a proper open subset V of U (Fig. 8). If P is in the closure of V but not in V, $u(P)$ and $u'(P)$ are distinct points of the sheaf but cannot be separated because every neighborhood of P contains points of V on which u and u' agree.

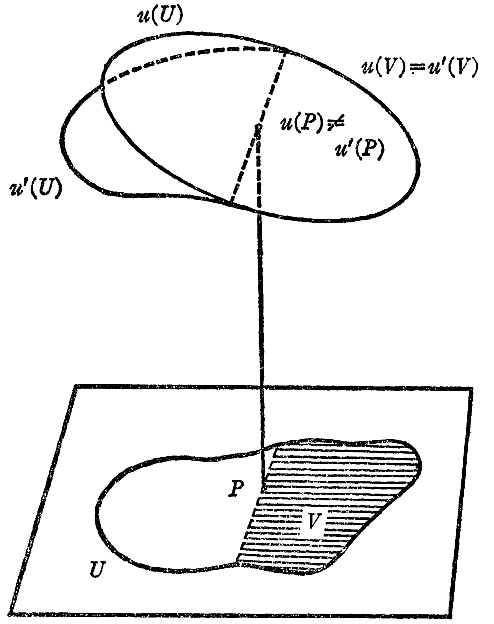


FIG. 8

A special case. In order to provide a more geometric interpretation of the sheaf of local rings as well as a glimpse of the origin of the subject, we shall look at some special rings. Let Q be the field of rational numbers, $Q[x_1, \cdots, x_n]$ the ring of polynomials in n variables with rational coefficients, and I an ideal of $Q[x_1, \cdots, x_n]$. Let A be the quotient ring $Q[x_1, \cdots, x_n]/I$, which we may write $Q[\bar{x}_1, \cdots, \bar{x}_n]$, where $\bar{x}_i = x_i + I$, $i = 1, \cdots, n$. If $E = \{(x_1, \cdots, x_n) \in C^n \mid p(x_1, \cdots, x_n) = 0$ for all $p \in I\}$, i.e., E is the intersection of the zero sets of all polynomials in the ideal I, we may interpret A as a ring of complex valued functions on E. For since $p_1(x) - p_2(x) \in I$, whenever $p_1(\bar{x}) = p_2(\bar{x})$, $p_1(\bar{x}) = p_2(\bar{x})$ implies $p_1(x) = p_2(x)$ if $x = (x_1, \cdots, x_n) \in E$.

We may also construct, from the pairing which takes $p \in A$ and $x \in E$ to $p(x) \in C$, a one-to-one correspondence between the points in E and homomorphisms from A to C. If $x = (x_1, \cdots, x_n) \in E$ the function which takes p to $p(x)$ is a homomorphism from A to C, and conversely, if $\phi: A \to C$ is a homomorphism, $x = (\phi(\bar{x}_1), \cdots, \phi(\bar{x}_n)) \in E$, since if $p \in I$,

$$p(x) = p(\phi(\bar{x}_1), \cdots, \phi(\bar{x}_n)) = \phi(p(\bar{x}_1), \cdots, p(\bar{x}_n)) = \phi(0) = 0.$$

Since C is an integral domain, the kernel of the homomorphism defined by $x \in E$ is a prime ideal in A. Conversely, it may be shown that if P is an element of Spec A it is the kernel of some homomorphism from A to C [13, p. 164]. So we may regard Spec A as a set of equivalence classes of E under the relation $x \sim x'$ if and only if x and x' are zeros of the same polynomials $p \in A$. Thus if $x \sim x'$, $p(x) = p(x')$ for all p in A, so A may also be interpreted as a ring of functions on Spec A.

That an element p of A not be in a prime P means that p is not in the kernel of the map $A \to A/P$, or that p is not zero on any point x in E for which the kernel of the corresponding homomorphism is contained in P. Thus, the local ring A_P at P, which contains the inverses for all $p \notin P$, is the ring of all rational functions defined locally, i.e., in some neighborhood of P in Spec A. The set of such functions which vanish at P is the only maximal ideal in this ring. The ring A_U of presheaf sections over an open set $U \subset \text{Spec } A$ is a ring of functions defined at every point of U. (If U is a distinguished open set, A_U is the ring of all such functions.) The kernel of a restriction map $\rho_{U,V}: A_V \to A_U$ (where $U \subset V$) is the set of functions which vanish at every point of U, though they may not vanish on all of V. The elements of A_U which are not restrictions of functions in A_V are inverses of functions which have zeros in $V - U$.

The topology of Spec A induces a topology in E, the weakest one in which the identification map from E to Spec A is continuous. In this topology, the closure of a point $x \in E$ is the set of all points in E which satisfy the same polynomials as x.

We choose a particular ring A to study in more detail. In $Q[x, y]$, let $I = (xy)$, and

$$A = Q[x, y]/(xy) \simeq Q[\bar{x}, \bar{y}],$$

where $\bar{x}\bar{y} = xy + (xy) = (xy) = 0$. Then

$$E = \{(z, w) \in C^2 \mid zw = 0\} = \{(z, 0) \mid z \in C\} \cup \{(0, w) \mid w \in C\},$$

the union of the complex coordinate axes (planes) in C^2.

The ideals $(\bar{x})$ and $(\bar{y})$ in A are prime but not maximal, since $Q[\bar{x}, \bar{y}]/(\bar{x}) \simeq Q[\bar{y}] \simeq Q[y]$, which is not a field. The points (z, w) in E whose corresponding ideal is $(\bar{x})$ are those of the form $(0, w)$ where w is transcendental over Q, and similarly, the points corresponding to the ideal $(\bar{y})$ are of the form $(z, 0)$ for transcendental z, since the homomorphism of $Q[\bar{x}, \bar{y}] \to C$ given by $\bar{x} \to 0$, $\bar{y} \to w$ has kernel $(\bar{x})$ if and only if w is transcendental. Further, $(\bar{x}) \cap (\bar{y}) = (0)$ in A, any prime ideal contains either $(\bar{x})$ or $(\bar{y})$, and the set of zero divisors in A is $(\bar{x}) \cup (\bar{y})$.

The maximal prime ideals in A are kernels of homomorphisms $\phi: [\bar{x}, \bar{y}] \to C$ for which the image is a field, so that $\phi(\bar{x})$ and $\phi(\bar{y})$ must be algebraic, and since $\phi(\bar{x})\phi(\bar{y}) = \phi(\bar{x}\bar{y}) = 0$, one of $\phi(\bar{x})$ and $\phi(\bar{y})$ must be zero, and the other algebraic. Such points $(\phi(\bar{x}), \phi(\bar{y}))$ in E are either of the form $(z, 0)$ with z algebraic, or $(0, w)$ with w algebraic. The kernels of the homomorphisms to

$(z, 0)$ and $(0, w)$ are the ideals $(p(\bar{x}), \bar{y})$ and $((\bar{x}), q(\bar{y}))$ where p is the minimal polynomial of z (and q of w). These ideals together with $(\bar{x})$ and $(\bar{y})$ are precisely the points of Spec A.

The points of $E \subset C^2$ are divided into equivalence classes corresponding to the ideals in Spec A: to each of (x) and (y) there corresponds a class of transcendental points, while to each maximal ideal $(p(\bar{x}), \bar{y})$ (where p is irreducible) there corresponds the set of algebraic points $\{(z, 0)\}$ where z is a root of p.

In the topology induced on E by Spec A, the closure of a transcendental point $(z, 0)$ is the z-axis and the closure of an algebraic point $(z, 0)$ is all $(z', 0)$ where z' is also a root of the minimal polynomial of z. A basis for the open sets in Spec A is the collection of all sets of the form $V_p = \{P \in \text{Spec } A \mid p \notin P\}$. In the special case $p = \bar{x}$,

$$V_{\bar{x}} = \{P \in \text{Spec } A \mid \bar{x} \notin P\} = \{(\bar{y}, p(\bar{x})) \mid p \in A \text{ is irreducible or zero, } p \neq \bar{x}\}.$$

The open set $V_{\bar{x}} \subset \text{Spec } A$ corresponds in E to the complement of the w-axis. Similarly, $V_{\bar{y}}$ may be regarded as the complement of the z-axis. The complement of a finite set of algebraic points corresponding to the maximal ideals

$$(\bar{x}, p_1(\bar{y})), \cdots, (\bar{x}, p_k(\bar{y})), \quad (p_{k+1}(\bar{x}), \bar{y}), \cdots, (p_n(\bar{x}), \bar{y})$$

is $V_{p_1(\bar{y})} \cap \cdots \cap V_{p_n(\bar{x})}$. Since every polynomial $p \in A = Q[\bar{x}, \bar{y}]$ can be written as $p = p_1(x) + p_2(y) - c$, where $p_1(0) = p_2(0) = p(0, 0) = c$, the points in V_p are $(z, 0)$ where $p_1(z) = 0$, and $(0, w)$ where $p_2(w) = 0$.

Geometrically, the ring $A = Q[\bar{x}, \bar{y}]$ is the ring of polynomials defined everywhere on E, the union of the complex axes. At a point (z, w) in E (where either z or w is 0) corresponding to a prime ideal P, the local ring A_P contains all rational functions whose denominators do not vanish at (z, w). The local ring $A_{\bar{x}}$ is the local ring at any $(0, w)$ where w is transcendental. Algebraically, we have the set of denominators $S = A - (\bar{x}) = \{p \mid p_2(\bar{y}) \neq 0\}$ where, as above, $p = p_1 + p_2 - c$. The kernel I of the homomorphism $\alpha: A \to A_{\bar{x}}$ is $\{q \in A \mid \exists p \in S$ such that $pq = 0\}$. If $p_1 + p_2 - c$ is a zero divisor, $c = 0$; if such a point is also in S, we have $p_2(\bar{y}) \neq 0$ and $p_2(0) = 0$, so it is an element of $(\bar{y}) - \{0\}$. Therefore $I = (\bar{x})$. Thus $A/I = Q[\bar{x}, \bar{y}]/(\bar{x}) = Q[\bar{y}]$ and the image of S in $Q[\bar{y}]$ is $Q[\bar{y}] - \{0\}$; so $A_{\bar{x}}$ is isomorphic to the field of rational functions in y. Similarly $A_{\bar{y}} = Q(x)$.

For the local ring at a maximal prime other than $(\bar{x}, \bar{y})$, for instance $P = (\bar{x}, \bar{y}^2 - 2)$, a similar argument will lead to the ring $A_P = \{p/q \in Q(y) \mid (y^2 - 2)$ does not divide $q\}$. That is, A_P contains inverses for all functions that do not vanish at $(0, \sqrt{2})$, while two polynomials in $\bar{x}$ and $\bar{y}$ which agree on an open set containing $(0, \sqrt{2})$ are identified in A_P. Thus the map $\alpha: A \to A_P$ is neither one-to-one nor onto. Now if $P = (\bar{x}, \bar{y})$, the set of denominators for A_P is exactly the set of polynomials $p_1 + p_2 - c$ for $c \neq 0$. So in this case the map $\alpha: A \to A_P$ is an inclusion which is not onto.

Finally, we describe the rings of presheaf sections A_U, for open sets U. If U is the complement of a finite closed set of maximal primes $(\bar{x}, p_1(\bar{y})), \cdots,$

$(\bar{x}, p_k(\bar{y}))$, $(p_{k+1}(\bar{x}), \bar{y})$, $\cdots$, $(p_n(\bar{x}), \bar{y})$ the set of denominators for A_U is $\{p \in A \mid p \notin \bigcup_{P \in U} P\}$, which is the set of all products $p_1^{\alpha_1} \cdots p_n^{\alpha_n}$ with nonzero constant term. The map from $A_U \to A_P$ when $P \in U$ will have kernel $(\bar{x})$ if $P = (\bar{x}, p(\bar{y}))$ (where p is any irreducible polynomial except $\bar{y}$); $(\bar{y})$ if $P = (\bar{y})$ or $(\bar{y}, p(\bar{x}))$ (where $p(x) \neq x$); and (0) if $P = (\bar{x}, \bar{y})$.

If $U = V_{\bar{x}}$ is the complement of the w-axis, the set of denominators for A_U is the set of polynomials $\bar{y}p(\bar{y}) + c\bar{x}^n$, for $p \in Q[\bar{y}]$ and $n \geqq 0$. Thus

$$A_U = \{p(x)/x^n \mid p \in Q[x]\}.$$

Since $A_{(\bar{y})} \simeq Q(x)$, the map $\rho_{(\bar{y}),U}: A_U \to A_{(\bar{y})}$ from the ring of sections to the stalk at $(\bar{y})$ is one-to-one but not onto. For a maximal prime $(\bar{y}, p) \in V_{\bar{x}}$, where $p(x) \neq x$, $\rho_{(\bar{y},p),\,U}: A_U \to A_{(\bar{y},p)}$ is an inclusion since x does not divide p. If U is the complement of the closed set $(\bar{x}, p_1(\bar{y}))$, $\cdots$, $(\bar{x}, p_k(\bar{y}))$, then $V_{\bar{x}} \subset U$, and the restriction map $\rho_{V_{\bar{x}},U}: A_U \to A_{V_{\bar{x}}}$ has kernel $(\bar{y})$.

The ring of sections over Spec A is A, for every element of A vanishes at some point in Spec A. Thus for every kind of open set U in Spec A, the elements of the rings A_U are the rational functions defined at every point of U.

In particular the functions $\bar{x}$ and $\bar{x} + \bar{y}$ are sections on the open set Spec A. They agree on the proper open subset $V_{\bar{x}}$, for $(\bar{y})$ is the kernel of the restriction map $\rho_{V_{\bar{x}},\,\mathrm{Spec}\,A}$, but do not agree on $(\bar{x}, \bar{y})$, which is in the closure of $V_{\bar{x}}$. Thus the points $\bar{x}$ and $\bar{x} + \bar{y}$ in $A_{(\bar{x},\bar{y})}$, though distinct, cannot be separated by sections since any open set containing $(\bar{x}, \bar{y})$ must intersect $V_{\bar{x}}$. Thus this sheaf fails to be Hausdorff both vertically as well as horizontally.

3. The sheaf of differential forms. Let $p \in X$ where X is a C^∞ manifold over R^n. We denote by C_p^∞ the set of all functions from X to the real line R^1 which are C^∞ in some neighborhood of p. Clearly C_p^∞ is a vector space over R in which the sum $f+g$ is defined on the intersection of the domains of f and g.

A *tangent* to X at p is a linear function $t: C_p^\infty \to R^1$ such that

$$t(fg) = t(g) \cdot g(p) + f(p) \cdot t(g) \qquad \text{whenever } f, g \in C_p^\infty.$$

The set of all tangents to X at p forms a vector space over R which we call X_p, the *tangent space to X at p*. If $\gamma: [0, 1] \to X$ is a C^∞ function such that $\gamma(t_0) = p$, and if $f \in C_p^\infty$, then $f \circ \gamma: [0, 1] \to X \to R^1$. If $(f \circ \gamma)'$ is the derivative of $f \circ \gamma$, the function $\gamma_*: C_p^\infty \to R^1$ defined by $\gamma_*(f) = (f \circ \gamma)'(t_0)$ is a tangent. To interpret this geometrically, we consider a particular local coordinate system (U_p, π_p) where $\pi_p: U_p \to R^n$. If $\pi_i: R^n \to R^1$ is the projection function defined by $\pi_i(t_1, \cdots, t_n) = t_i$, we write $x_i = \pi_i \circ \pi_p$; then $\pi_p(p) = (x_1(p), \cdots, x_n(p)) \in R^n$. If $e_i = (0, 0, \cdots, 1, \cdots, 0) \in R^n$, we think of the curve $\gamma(t) = \pi_p^{-1}(\pi_p(p) + te_i)$ as the ith coordinate axis in U_p since $(d/dt)(\pi_p \circ \gamma)\big|_0 = e_i$. Then the tangent γ_* satisfies

$$\gamma_*(f) = (f \circ \gamma)'(0) = \frac{d}{dt}(f \circ \pi_p^{-1} \circ \pi_p \circ \gamma)\bigg|_0$$

$$=\left(\frac{\partial(f\circ\pi_p^{-1})}{\partial t_1},\cdots,\frac{\partial(f\circ\pi_p^{-1})}{\partial t_n}\right)\Bigg|_{\pi_p(p)}\cdot e_i=\frac{\partial(f\circ\pi_p^{-1})}{\partial t_i}\Bigg|_{\pi_p(p)}.$$

Thus $\gamma_*(f)$ is usually denoted by $(\partial/\partial x_i)f$; the tangents $(\partial/\partial x_i)$, $\cdots$, $(\partial/\partial x_n)$ form a basis for the vector space X_p [8, p. 7]. We think of X_p as an n-dimensional hyperplane tangent to X at p (Fig. 9). If

$$\gamma_*=\sum a_i\frac{\partial}{\partial x_i},$$

then $\gamma_*(f)$ is $a\cdot\operatorname{grad} f$, that is, the derivative of f in the direction $a=(a_1,\cdots,a_n)\in R^n$.

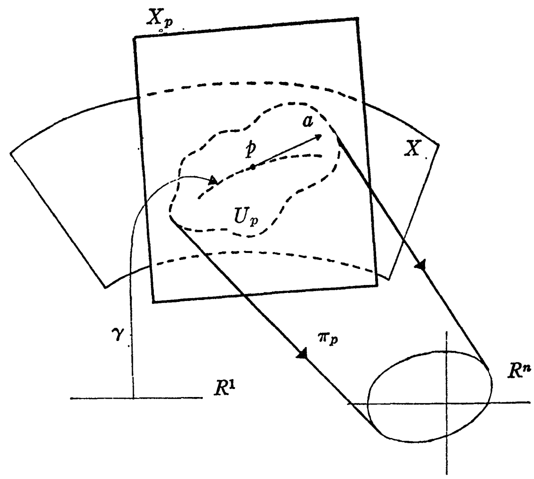


FIG. 9

If $\phi:X\to Y$ is a C^∞ mapping of manifolds we define the *differential of* ϕ to be the linear map $d\phi:X_p\to Y_{\phi(p)}$ defined as follows: if $t\in X_p$ and $f\in C^\infty_{\phi(p)}$ then $[d\phi(t)](f)=t(f\circ\phi)$. If (U_p,π_p) is a local coordinate system at $p\in X$ then $x_i=\pi_i\circ\pi_x$ defines a C^∞-map from the submanifold U_p to the manifold R^1. Hence the differential dx_i of x_i is a linear transformation from X_p to $R^1_{x_i(p)}$. Since R^1_t is isomorphic to R^1 for any $t\in R^1$, we may consider dx_i as an element of the dual vector space X_p^* of X_p. Since $dx_i\,(\partial/\partial x_j)=\delta_{ij}$ (where $\delta_{ij}=0$ if $i\neq j$, and 1 if $i=j$), the differentials dx_i, $\cdots$, dx_n form a basis of X_p^* dual to the basis $\partial/\partial x_i$, $\cdots$, $\partial/\partial x_n$ of X_p.

Whereas each differential dx_i is a function of just one variable, a differential form in general is a function of several variables. To be precise, a *differential k-form θ* on X_p is an alternating k-linear function from k-tuples of elements of X_p to R^1. (θ is *k-linear* if it is linear in each variable separately, and *alternating* if $f(t_1\cdots t_k)=\operatorname{sgn}\sigma f(t_{\sigma(1)}\cdots t_{\sigma(k)})$ where $\sigma:\{1\cdots k\}\to\{1\cdots k\}$ is a permutation and $\operatorname{sgn}\sigma$ is $+1$ if σ is even, and -1 if σ is odd.) We denote by the *wedge product* $dx_{i_1}\wedge\cdots\wedge dx_{i_k}$ the unique k-linear form on X_p defined on a basis for

the set of k-tuples by

$$dx_{i_1} \wedge \cdots \wedge dx_{i_k}\left(\frac{\partial}{\partial x_{j_1}}, \cdots, \frac{\partial}{\partial x_{j_k}}\right) = \delta,$$

where $\delta = 1$ if $(i_1, \cdots, i_k) = (j_1, \cdots, j_k)$ and 0 otherwise. The set of all k-linear forms on X_p is an n^k dimensional vector space over R and the n^k forms $dx_{i_1} \wedge \cdots \wedge dx_{i_k}$, where $i_j \in \{1, \cdots, n\}$, form a basis for this vector space. The set of differential k-forms is a subspace of the set of all k-linear forms and it has as a basis the k-linear forms $dx_{i_1} \wedge \cdots \wedge dx_{i_k}$ where $i_1 < i_2 < \cdots < i_k$. We call such a sequence of indices an *increasing k-tuple*. Thus the dimension of the space of differential k-forms is $\binom{n}{k}$ if $k \leqq n$ and 0 if $k > n$.

Let S_k denote the set of all increasing k-tuples of positive integers less than or equal to n. For $x \in S_k$ we denote by dx_s the k-form $dx_{i_1} \wedge \cdots \wedge dx_{i_k}$, where $s = (i_1, \cdots, i_k)$. Now let $U \subset X$ be an open set and θ a function which assigns to each $p \in U$ a differential k-form at p, $\theta(p)$. Thus, once having chosen a local coordinate system at p, so the forms $dx_1, dx_2, \cdots, dx_n$ are defined, we may write

$$\theta(p) = \sum_{s \in S_k} a_s(p) dx_s,$$

where each $a_s(p)$ is the coefficient in the expansion of $\theta(p)$ in terms of the basis $\{dx_s\}_{s \in S_k}$. Thus a_s may be considered as a function from U to R^1. Since each $\theta(p)$ is a differential k-form at p, we call θ a C^∞ *k-form* on U if each function a_s is C^∞; we denote by $\Omega^k(U)$ the real vector space of all C^∞ k-forms on U. If $U \subset V$ then there is a linear transformation $\rho_{U,V}: \Omega^k(V) \to \Omega^k(U)$ defined by restriction of the domains of $\theta \in \Omega^k(V)$. The collection $\{\Omega^k(U)\}$ for U open in X together with the linear transformations $\rho_{U,V}$ is called the *presheaf of differential k-forms*.

Since every paracompact manifold has a C^∞ partition of unity subordinate to any open covering $\{U_\alpha\}$ [8, p. 85], presheaves over such manifolds satisfy the following special property: If $U = \bigcup U_\alpha$ where U_α is open in X, and if $\theta_\alpha \in \Omega^k(U_\alpha)$ are coherent in the sense that the restrictions to $U_\alpha \cap U_\beta$ of θ_α and θ_β agree (whenever $U_\alpha \cap U_\beta \neq \emptyset$) then there exists a unique $\theta \in \Omega^k(U)$ whose restriction to each U_α is θ_α. Certainly if $\{f_\alpha\}$ is a C^∞ partition of unity for $\{U_\alpha\}$ so that $f_\alpha: U \to R^1$, f_α vanishes off U_α, and $\sum f_\alpha = 1$; so we may define θ to be $\sum f_\alpha \theta_\alpha$.

Any such presheaf is called a sheaf, so when X is paracompact, we will call the system $\{\Omega^k(U), \rho_{U,V}\}$ the *sheaf of differential k-forms*. If $\Omega^k(p)$ is the set of differential k-forms at p, we may think of $\Omega^k(p)$ as the stalks of the sheaf S: $S = \bigcup_{p \in X} \Omega^k(p)$. The projection $\rho: S \to X$ assigns to each differential form at p the point p. The C^∞ k-form $\theta \in \Omega^k(U)$ is a section of S, and the collection $\{\theta(U) \subset S \mid U \text{ is open in } X\}$ forms a basis for the topology on S.

4. Sheaves: General definition. Each of the three previous examples reflects a different facet of the general concept of a sheaf. To emphasize this

	Sheaf of Germs of Holomorphic Functions	Sheaf of Local Rings	Sheaf of Differential Forms
Germ	$[f]_z$	$a/s \in A_P$	
Stalk Space	$S=\{(x, [f]_x)\}$	$S=\bigcup\{A_P \mid P\in \text{Spec } A\}$	
Projection	$\rho: S\to C^n$	$\rho: S\to \text{Spec } A$	
Stalk	$\{(z, [f]_z) \mid f\in A_z\}$	A_P(local ring)	
Base Space	C^n	Spec $A=\{P \mid P$ is prime ideal in $A\}$	$X=C^\infty$manifold
Presheaf Section		$a/s\in A_U$	$\theta\in\Omega^k(U)$
Restriction Homomorphism		$\rho_{U,V}: A_V\to A_U$	$\rho_{U,V}: \Omega^k(V)\to\Omega^k(U)$

FIG. 10

variety and to provide a coherent framework for the subsequent general definition, we summarize in Fig. 10 the three sheaves already discussed.

The sheaf of germs of holomorphic functions was defined to be the stalk space S together with a topology and a local homeomorphism ρ onto C^n. The sheaf of local rings was defined similarly, though in this case we also identified the system consisting of the rings A_U and the restriction homomorphisms $\rho_{U,V}: A_V\to A_U$ as a presheaf of rings. In the third case, the sheaf of differential k-forms was simply the presheaf of differential k-forms whenever the base space X was paracompact. The recognition of the equivalence of these two descriptions constitutes the beginning of sheaf theory. We now introduce definitions to formalize these two approaches and prove the definitions equivalent.

DEFINITION I. Let (X, τ) be a topological space, and let $\mathcal{C}$ be a class of similar mathematical objects (e.g., abelian groups, modules, rings). Let F be a function from τ to $\mathcal{C}$ and suppose for each pair U, $V\in\tau$ for which $U\subset V$ there is a map (e.g., a homomorphism, module homomorphism, or isomorphism) $\rho_{U,V}: F(V)\to F(U)$ which preserves the structure of the objects of $\mathcal{C}$. If $\rho_{U,V}\circ\rho_{V,W}=\rho_{U,W}$ whenever $U\subset V\subset W$, and if $\rho_{U,U}$ is the identity function on $F(U)$, we define the function F together with the *restriction* maps $\rho_{U,V}$ to be a *presheaf* over X. The elements of $F(U)$ are called *sections* of F over U. (In the language of category theory, a presheaf is a contravariant functor from the category of open sets and inclusion maps of X to some category of objects and morphisms.) A *sheaf* is a presheaf which satisfies the following two coherence axioms:

1. *If $\{U_\alpha\}$ is a family of open sets in X, if $U=\bigcup U_\alpha$, and if the section s, $t\in F(U)$ agree on each U_α (i.e., if $\rho_{U_\alpha,U}(s)=\rho_{U_\alpha,U}(t)$ for each α), then $s=t$.*

2. *If $\{U_\alpha\}$ is a family of open sets in X, if $U=\bigcup U_\alpha$,* and if the sections *$s_\alpha\in F(U_\alpha)$ are coherent in the sense that the restrictions of s_α and s_β to $U_\alpha\cap U_\beta$ agree (i.e., if $\rho_{U_\alpha\cap U_\beta,U_\alpha}(s_\alpha)=\rho_{U_\alpha\cap U_\beta,U_\beta}(s_\beta)$ whenever $U_\alpha\cap U_\beta\neq\emptyset$) then there exists a section $s\in F(U)$ such that for each α, $\rho_{U_\alpha,U}(s)=s_\alpha$.*

DEFINITION II. Let (X, τ) be a topological space, and let $\mathcal{C}$ be a class of similar mathematical objects (e.g., abelian groups, modules, rings). A *sheaf* over X is a triple $\{S, \pi, X\}$, where S is a topological space and $\pi: S \to X$ is a local homeomorphism (i.e., a map such that each point $p \in S$ has a neighborhood U on which $\pi|_U$ is a homeomorphism) such that each *stalk* $\pi^{-1}(x) \in \mathcal{C}$, and each operation is continuous as a function from $\bigcup_{x \in X}(\pi^{-1}(x) \times \pi^{-1}(x))$ (with the topology induced from $S \times S$) to $\bigcup_{x \in X} \pi^{-1}(x) = S$.

We shall call a sheaf of type I a *sheaf of sections*, and a sheaf of type II a *sheaf of germs*. The relationship between these two types of sheaves is precisely as illustrated by the preceding examples.

To be specific, suppose F is a presheaf (of sections) over W; we construct the corresponding sheaf of germs by defining a germ at $x \in X$ to be an equivalence class of $A_x = \bigcup_{x \in U} F(U)$ under the relation $s \in F(U) \sim t \in F(V)$ if $\rho_{W,U}(s) = \rho_{W,V}(t)$ for some $W \subset U \cap V$. Thus the germ of a section $s \in F(U)$ at a point $x \in U$ is the collection of all sections $t \in F(V)$ which agree with s on some neighborhood V of x. We denote, as usual, the germ of s at x by $[s]_x$, and let the stalk space S be $\{(x, [s]_x) | s \in F(U)$, where $x \in U\}$. The topology on S is generated by neighborhoods of the form $V(s, U) = \{(x, [s]_x | x \in U\}$, so the projection $\pi: S \to X$ becomes a local homeomorphism. By interpreting the sections as functions, the so-called restriction maps $\rho_{U,V}$ really are restrictions and the topology on S is the strongest relative to which the sections are continuous. (The topology on S can also be characterized as the quotient of the topology on $\bigcup_{U \in \tau} (U \times F(U))$ under the equivalence relation induced by $\sim$, where each U carries the subspace topology and $F(U)$ is discrete.)

Each stalk $\pi^{-1}(x)$ clearly inherits the operations of the $F(U)$ and each such operation is continuous. For example, if each $F(U)$ is an abelian group under addition and if $[s]_x$ and $[t]_x \in \pi^{-1}(x)$, then $[s]_x + [t]_x$ is defined to be $[s+t]_x$. If $V(s+t, U)$ is a neighborhood of $[s+t]_x$, then the inverse image of $V(s+t, U)$ under $+$ contains $\{(r, q) | r \in V(s, U), q \in V(t, U), \pi(r) = \pi(q)\}$, an open set in $\bigcup_{x \in X}(\pi^{-1}(x) \times \pi^{-1}(x))$. Thus $\{S, \pi, X\}$ is indeed a sheaf of germs.

Conversely, suppose $\{S, \pi, X\}$ is a sheaf of germs (perhaps one constructed as above from some presheaf). If U is open in X, let $F(U)$ be the collection of continuous functions $s: U \to F$ such that $\pi \circ s$ is the identity on U. $F(U)$ inherits the algebraic structure from the stalks by pointwise definitions, and the restriction maps $\rho_{U,V}$ are just that—the restriction of s from V to the subset U. The first coherence axiom for sheaves is satisfied trivially since the sections are functions, and the second is satisfied since the $F(U)$ contain all continuous functions from U to F which are inverses of π.

Now if $\{F, \pi, X\}$ is a sheaf of germs, the sheaf derived from its (pre)sheaf of sections is canonically isomorphic to F. However, if S is a presheaf of sections, the presheaf of sections S' associated with the sheaf of germs derived from S is generally different from S, for since S' is a sheaf, it may have more sections than S (in order to satisfy the second coherence axiom), while some sections which were distinct in S may be identified in S' (because of the first axiom). Of course,

if S is a sheaf, then S' is naturally isomorphic to S, so in this sense the two definitions of a sheaf are essentially equivalent.

A common and convenient alternative to the construction of equivalence classes is the use of direct limits. If F is a presheaf of sections over X, and if $x \in X$, the restriction of F to the neighborhoods of x forms a *directed system* $(\{F(U)\}_{x\in U}, \{\rho_{U,V}\}_{x\in U\subset V})$ since $\rho_{U,V} \circ \rho_{V,W} = \rho_{U,W}$, whenever $x \in U \subset V \subset W$. A few elements of such a system may be represented by the commutative diagram of Fig. 11.

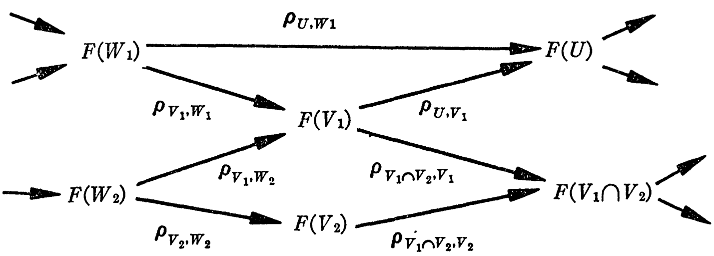


FIG. 11

The direct limit of this system is, roughly speaking, the first object which can appear to the right of the diagram. Specifically, an object F_x together with maps $\rho_U: F(U) \to F_x$ for each $F(U)$ is called a *direct limit* of the system $(\{F(U)\}_{x\in U}, \{\rho_{U,V}\}_{x\in U\subset V})$ provided that

(i) whenever $U \subset V$ the diagram in Fig. 12 commutes,

(ii) F_x is universal with respect to property (i)—that is, if $(G_x, \{\sigma_v\})$ also satisfies property (i), then there exists a unique map $\eta: F_x \to G_x$ such that for each $\sigma_U: F(U) \to G_x$, $\sigma_U = \eta \circ \rho_U$.

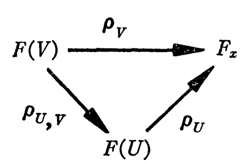


FIG. 12

Condition (i) makes explicit the idea of "appearing on the right of the diagram" while condition (ii) asserts that F_x is the first such object. It follows trivially from these conditions that the direct limit is unique (up to isomorphism), so we denote it by $\varinjlim_{x\in U} F(U)$.

Consider now the stalk of germs at x derived from the presheaf F. If ρ_U denotes the map from $F(U)$ to the stalk $\pi^{-1}(x)$ defined by $\rho_U(s) = (x, [s]_x)$ (where $x \in U$), then $\pi^{-1}(x) = \varinjlim_{x\in U} F(U)$ since whenever $x \in U \subset V$, $\rho_V = \rho_U \circ \rho_{U,V}$ and $\pi^{-1}(x)$ is universal with respect to that property. To prove this last assertion we assume that $(G, \{\sigma_U\})$ is another direct limit and observe that if both

$t_1 \in F(U_1)$ and $t_2 \in F(U_2)$ are in the same germ $[s]_x$, then $\sigma_{U_1}(t_1) = \sigma_{U_2}(t_2)$, for by the definition of $[s]_x$, there exists some $V \subset U_1 \cap U_2$ such that $\rho_{V,U_1}(t_1) = \rho_{V,U_2}(t_2)$. Then by the commutativity of the diagram in Fig. 13, we have

$$\sigma_{U_1}(t_1) = \sigma_V \circ \rho_{V,U_1}(t_1) = \sigma_V \circ \rho_{V,U_2}(t_2) = \sigma_{U_2}(t_2).$$

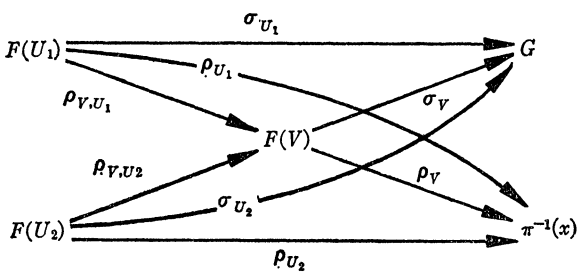


FIG. 13

Thus each element $[s]_x$ of $\pi^{-1}(x)$ is mapped to some point of G by the appropriate $\sigma_U \circ \rho_U^{-1}$: we call this point $\eta(x, [s]_x)$ and thereby define the required unique map from $\pi^{-1}(x)$ to G.

So we may summarize all three approaches in one sweeping generalization: for each $x \in X$, the stalk over x of the sheaf of germs is the direct limit of the restriction of the presheaf to the neighborhoods of x.

We close this section with a fourth characterization of sheaves, this one also based on a universal property, and illustrated to some extent in the previous examples. Suppose $\mathfrak{F}$ is a class of functions defined on open subsets of a topological space X. If $F(U)$ is the collection of all $f \in \mathfrak{F}$ whose domain is U, and if $\rho_{U,V}$ is the restriction map (i.e., $\rho_{U,V}(f) = f|_U$ whenever $U \subset V$ and $f \in F(V)$), the collection $\{F(U), \rho_{U,V}\}$ is a presheaf. If this presheaf is a sheaf S (as it will be if $\mathfrak{F}$ is the set of holomorphic functions on open subsets of C^n, or the set of differential forms on open subsets of a paracompact manifold X) then we can define a universal continuous function $\Phi: S \to R$ of type $\mathfrak{F}$ so that each $f \in F(U)$ factors uniquely through the sheaf S: that is, there exists a unique $\hat{f}: U \to S$ such that $\hat{f} = \Phi \circ f$.

In general, any object S and map $\Phi: S \to R$ with the property that each function $f: U \to R$ factors uniquely through S via Φ is called universal with respect to the characterizing properties of the functions f. The pair (S, Φ) is uniquely determined (up to isomorphism) by this property. Thus, for instance, the sheaf of holomorphic functions is characterized by being the unique universal object for the family of holomorphic functions.

Since $\hat{f}^{-1}(V(f, U)) = U$, $\hat{f}$ is continuous. Thus sheaves transform a complicated property of functions, such as analyticity, into the simpler one of continuity, for the topology on the sheaf is chosen precisely so that a continuous section on U (i.e., an element of $F(U)$) corresponds to one of the specialized (e.g., analytic, differentiable) functions of $F(U)$.

5. History and applications. Sheaf theory is a particularly effective tool in those areas which ask for global solutions to problems whose hypotheses are local. Among the early papers which introduced the ideas, though not the language, of sheaf theory, many were concerned with the Cousin problems from the theory of functions of several complex variables; the first (or additive) and second Cousin problems ask respectively about the existence of a meromorphic function with specified poles and the existence of a holomorphic function with specified zeros. Henri Cartan and Kiyoshi Oka independently solved these problems, working in the ring of germs of holomorphic functions introduced in our first example, where the operations take into account the domains of the functions. Oka [1950] cites Cartan [1940] as the source of the notion of "*idéal holomorphe de domaines indéterminés*" in this ring, and both Oka [1951] and Cartan [1944] refer to the article of W. Rückert [1933] which took the concept of ideal from polynomial rings and interpreted it in the ring of functions on a fixed domain. Cartan [1944] carried on the investigation of the sheaf of germs, still in the earlier terminology, clarifying the relations among the problems without achieving solutions.

Independently, Oka in 1948 wrote a paper [1950] (seventh in a series published from 1936 to 1953 and collected in a single volume [1961]) which developed the same material in a more complete form, and carried it through to a solution of the first Cousin problem. Building on Oka's paper, Cartan was able to solve the second problem as well as to simplify Oka's solution to the first, and his paper [1950] and Oka's were published together. A footnote acknowledges Oka's solution of the second problem in the meantime [1951].

The 1950 Cartan paper for the first time phrases the questions in the sheaf theoretic terms which had been developed in the *Séminaire Cartan* in 1948–49. An analytic sheaf, that is a sheaf of modules over the sheaf of germs of holomorphic functions, is called *coherent* over an open set U if for every $x \in U$ there is an open set U_x such that the sections over U_x generate the stalk at y for all y in a sufficiently small neighborhood of x. If $f_1, \cdots, f_k$ are functions holomorphic on a domain D, we may define the sheaf R of relations among the f_i by taking the sections R_U of R over U open in D to be the set of k-tuples of holomorphic functions $(g_1, \cdots, g_k)$ for which $\sum_{i=1}^{k} f_i g_i \equiv 0$ on U. In this vocabulary the first Cousin problem is to show that R is coherent, while the second problem similarly asks whether the sheaf over an analytic variety is coherent, where a variety is the set of common zeros of a set of holomorphic functions, and the ideal of sections over an open set is the ideal of functions on the variety which vanish on that open set.

Cartan borrowed the term "faisceau" (sheaf) from Leray [1945; 1946]. Leray's concept was closer to that of a "presheaf." Cartan [1953] attributes the topological definition to an exposition by Lazard in the *Séminaire Cartan* [1950]. In each case, the key concept was that of a system of local coefficients. Studying sets of invariants for an object (base space) by investigating what functions can be defined from it to some convenient object called a set of co-

efficients, as is done in cohomology, leads very naturally to a sheaf of coefficients since the presheaf structure allows coefficients to be assigned locally, that is, to each open subset of the base space. Formally, the principal construction of cohomology with coefficients in a (pre)sheaf follows the Čech construction of cohomology with fixed coefficients.

Let X be a topological space and S a sheaf over X, say of abelian groups. For any open cover $\mathfrak{U}$ of X, a *q-cochain*, q being a nonnegative integer, is an alternating function which assigns to every $q+1$-tuple of sets in the cover $\mathfrak{U}$ a section over the intersection of these sets (the zero section if the intersection is empty). $C^q(\mathfrak{U}, S)$ denotes the group of q-cochains. For each q a *coboundary operator* δ^q, $\delta^q: C^q(\mathfrak{U}, S) \to C^{q+1}(\mathfrak{U}, S)$ is defined by

$$\delta^q f(U_{i_0}, \cdots, U_{i_{q+1}}) = \sum_{j=0}^{q+1} (-1)^j f(U_{i_0}, \cdots, \hat{U}_{i_j}, \cdots, U_{i_{q+1}}),$$

where the caret over U_{i_j} means that U_{i_j} is to be omitted from the arguments of f, and each of the sections on the right is to be interpreted as restricted to the intersection of all the U_{i_j}. By convention we write δ for all δ^q. Since f is alternating, $\delta\delta = 0$, so the image $\delta(C^{q-1}(\mathfrak{U}, S))$, whose elements are called *coboundaries*, is contained not merely in $C^q(\mathfrak{U}, S)$, but in the set of *cocycles* $Z^q(\mathfrak{U}, S)$, the kernel of

$$\delta: C^q(\mathfrak{U}, S) \to C^{q+1}(\mathfrak{U}, S).$$

The qth *cohomology group* $H^q(\mathfrak{U}, S)$ of the cover $\mathfrak{U}$ with coefficients in S is the quotient group $Z^q(\mathfrak{U}, S)/\delta(C^{q-1}(\mathfrak{U}, S))$. Although the construction of the cohomology group uses only the presheaf of sections of S, the sheaf property allows us to interpret $H^0(\mathfrak{U}, S)$. In order for 0-cochain to be a cocycle,

$$\delta f(U_0, U_1) = \rho_{U_0 \cap U_1, U_1}(f(U_1)) - \rho_{U_0 \cap U_1, U_0}(f(U_0))$$

must be zero, and in any sheaf, a collection of sections so related defines a unique global section. Thus $H^0(\mathfrak{U}, S)$ is S_X—independent of the cover $\mathfrak{U}$. For all q, if $\mathfrak{U}$ is a covering which refines a covering $\mathfrak{V}$, the restriction maps can be used to define a canonical map $H^q(\mathfrak{V}, S) \to H^q(\mathfrak{U}, S)$. The direct limit, over all coverings $\mathfrak{U}$ of X, of the groups $H^q(\mathfrak{U}, S)$ with these maps, is the qth *cohomology group* of X with coefficients in the sheaf S and is denoted by $H^q(X, S)$.

One property frequently taken as axiomatic for cohomology theories holds also for this one. If $0 \to F \xrightarrow{f} G \xrightarrow{g} H \to 0$ is an exact sequence of sheaves over X, there is a *long exact sequence* of cohomology

$$\begin{aligned} 0 \to H^0(X, F) \to H^0(X, G) \to H^0(X, H) \to H^1(X, F) \to \cdots \\ \to H^{q-1}(X, H) \to H^q(X, F) \to H^q(X, G) \to H^q(X, H) \to H^{q+1}(X, F) \to \cdots \end{aligned}$$

[2, p. 28]. As usual, a pair of homomorphisms $A \xrightarrow{f} B \xrightarrow{g} C$ is *exact* at B if $\operatorname{Im} f = \operatorname{Ker} g$, so the exactness of $0 \to F \xrightarrow{f} G \xrightarrow{g} H \to 0$ means that F is a subsheaf of G (this requires f to be an open mapping), g is onto, and H is isomorphic to the

quotient sheaf G/F. The existence of the long exact sequence is the major reason for the usefulness of cohomology, for the 0-dimensional groups which begin the sequence are the groups of global sections, while the higher groups by their construction reflect the local properties of X.

For instance, if we take F to be the sheaf of germs of holomorphic functions on a complex manifold X, and G to be the sheaf of germs of meromorphic functions, then F is a subsheaf of G, and the global sections of the quotient sheaf G/F can be interpreted as the data of the first Cousin problem, since each section describes the behavior of a function near its poles [**9**, p. 161]. Thus, since this problem asks whether there exists a function meromorphic on X with such poles, the first Cousin problem may be interpreted as asking whether the last map in the sequence $0 \to F_X \to G_X \to (G/F)_X$ is onto. This sequence is the beginning of the long exact cohomology sequence, and the next group in that sequence is $H^1(X, F)$. The Cartan-Oka result is that $H^q(X, F) = 0$ for all $q \geqq 1$ if X is a Stein manifold, a class of manifolds with "sufficiently many" holomorphic functions, which includes all Riemann surfaces which are connected and non-compact. In addition to proving this result, Cartan [1953] and Serre [1953] give other applications of the fundamental theorems for a Stein manifold X:

THEOREM A. *For every coherent analytic sheaf S over X, $H^0(X, S)$, which is the module of global cross sections S_X, generates the stalk S_x for every $x \in X$.*

THEOREM B. *For every coherent analytic sheaf S over X and $q \geqq 1$,*

$$H^q(X, S) = 0.$$

Properties A and B characterize Stein manifolds.

The proceedings of the 1954 AMS summer institute [1956] illustrate that by then the basic concepts of sheaf theory had been clarified apart from the original example, and the bibliographies indicate that applications had begun to diversify, particularly into algebraic geometry. For example, Kodaira and Spencer showed the equivalence of several different definitions of the arithmetic genus of an algebraic variety and provided a classification of complex line bundles [1953]. Hirzebruch gave a sheaf-theoretic statement and proof of the Riemann-Roch theorem [1953; 1956], and Weil of the deRham theorem [1952].

One seminar at the 1954 institute was based on an early version of Serre's major article FAC [1955], the first entirely algebraic development of sheaf theory. The applications to complex variables had frequently made use of complex integration, and this tool was not available in abstract algebraic geometry. "Faisceaux algébriques cohérents" are coherent sheaves which are sheaves of modules over the sheaf of local rings on an algebraic variety. Serre showed that if the base field is the field of complex numbers, the theory of algebraic coherent sheaves is isomorphic to the theory of analytic coherent sheaves. Going further in the direction of an algebraic treatment, Grothendieck dealt with sheaves in the context of cohomology in an abelian category [1957].

The publication of Godement's book [1958] signals the appearance of sheaf theory as an independent discipline. The bibliography which follows lists both recent treatments of sheaf theory and books on other subjects which make use of sheaves.

Historical Bibliography

1933. W. Rückert, Zum Eliminationsproblem der Potenzreihenideale, *Mathematische Annalen,* 107 (1933) 259–281.

1940. Henri Cartan, Sur les matrices holomorphes de n variables complexes, *Journal de Mathématiques Pures et Appliquées,* 19 (1940) 1–26.

1944. ———, Idéaux de fonctions analytiques de n variables complexes, *Annales de l'École Normale,* 3e série, 61 (1944) 149–197.

1945. Jean Leray, Sur la forme des espaces topologiques . . . (3 articles), *Journal de Mathématiques Pures et Appliquées,* 24 (1945) 95–248.

1946. ———, L'anneau d'homologie d'une représentation, *C. R. Acad. Sci.,* 222 (1946) 1366–1368.

1950. Kiyoshi Oka, Sur quelques notions arithmétiques, *Bulletin de la Société Mathématique de France,* 78 (1950) 1–27.

1950. Henri Cartan, Idéaux et modules de fonctions analytiques de variables complexes, *Bulletin de la Société Mathématique de France,* 78 (1950) 29–64.

1950. *Séminaire Cartan.* 1950–51. vol. 1, Benjamin, New York, 1967.

1951. Kiyoshi Oka, Lemme fondamental, *Journal of the Mathematical Society of Japan,* 3 (1951) 204–214, 259–278.

1952. André Weil, Sur les théorèmes de deRham, *Commentarii Mathematici Helvetici,* (1952) 119–146.

1953. Henri Cartan, Variétés analytiques complexes et cohomologie, *Colloque de Bruxelles,* (1953) 41–55.

1953. K. Kodaira and D. C. Spencer, On Arithmetic Genera of Algebraic Varieties, *Proceedings of the National Academy of Sciences USA,* 39 (1953) 641–649.

1953. ———and———, Groups of complex line bundles over compact Kähler varieties, *Proceedings of the National Academy of Sciences USA,* 39, (1953) 868–877.

1953. Jean-Pierre Serre, Quelques problèmes globaux relatifs aux variétés de Stein, *Colloque de Bruxelles,* (1953) 57–68.

1954. F. Hirzebruch, Arithmetic genera and the theorem of Riemann-Roch for algebraic varieties, *Proceedings of the National Academy of Sciences USA,* 40 (1954) 110–114.

1955. Jean-Pierre Serre, Faisceaux algébriques cohérents, *Annals of Mathematics,* 61 (1955) 197–278.

1956. F. Hirzebruch, *Topological Methods in Algebraic Geometry,* 3rd ed., Springer, Berlin, 1966.

1956. Shiing-Shen Chern, Scientific Report of the Second Summer Institute; II, Complex Manifolds, *Bulletin, American Mathematical Society,* 62 (1956) 102–117.

1956. Oscar Zariski, Scientific Report of the Second Summer Institute; III, Algebraic Sheaf Theory, *Bulletin, American Mathematical Society,* 62 (1956) 117–141.

1957. Alexandre Grothendieck, Sur quelques points d'algèbre homologique, *Tôhoku Mathematical Journal* Ser 2, 9 (1957) 119–221.

1958. Roger Godement, *Topologie Algébrique et Théorie des Faisceaux,* Hermann, Paris, 1958.

1961. Kiyoshi Oka, *Sur les Fonctions Analytiques de plusieurs variables,* Iwanami Shoten, Tokyo, 1961.

General Bibliography

1. Lipman Bers, Introduction to Several Complex Variables, Courant Institute, New York, 1964.

2. Glen Bredon, Sheaf Theory, McGraw-Hill, New York, 1967.

3. C. H. Dowker, Lectures on Sheaf Theory, Tata Institute of Fundamental Research, Bombay, 1956.
4. J. W. Gray, Sheaves with values in a category, Topology, 3 (1965) 1–18.
5. Alexandre Grothendieck and Jean Dieudonné, Éléments de Géométrie Algébrique, Publications Mathématiques de l'Institut des Hautes Études Scientifiques, Nos. 4, 7, 11, 17, 20, 24, 28, 32.
6. Robert Gunning and Hugo Rossi, Analytic Functions of Several Complex Variables, Prentice Hall, Englewood Cliffs, N. J., 1965.
7. M. Hervé, Several Complex Variables, Oxford Univ. Press, 1963.
8. Noel J. Hicks, Notes on Differential Geometry, Van Nostrand, Princeton, N. J., 1965.
9. Lars Hörmander, Complex Analysis in Several Variables, Van Nostrand, Princeton, N. J., 1966.
10. Ian Macdonald, Algebraic Geometry, Introduction to Schemes, Benjamin, New York, 1968.
11. Richard Swan, The Theory of Sheaves, University of Chicago Press, Chicago, 1964.
12. O. Zariski and P. Samuel, Commutative Algebra I, Van Nostrand, 1958.
13. ——— and ———, Commutative Algebra II, Van Nostrand, 1960.

ON THE FOUNDATIONS OF SET THEORY

J. D. MONK, University of Colorado

I want to discuss here the relevance to mathematicians, as teachers and researchers, of some of the recent discoveries about axiomatic set theory. Most readers have heard of these advances, which began just a few years ago with Cohen's work. The results are certainly intellectually amazing to all of us. I think they may even give rise to certain changes in our teaching and research, and the purpose of this paper is to describe some possibilities along these lines. To set the stage and fix the ideas I shall first describe a few of these discoveries in a fairly precise way. Then, in the nonexact portion of the paper, I shall discuss some possible changes in teaching and research, and also some philosophical views which are affected by these discoveries.

1. A survey of results. A much more comprehensive (and more technical) survey can be found in Mathias [7]. Here I state just a very few results, but I wish to emphasize that the nonmathematical arguments of the next section apply in some form to virtually all of the results described in [7]. I assume that the reader has a modest acquaintance with the idea of a language and a metalanguage, and with the precise notions of a (first-order) sentence, a (formal) proof, and a theorem. In this section I work in a metalanguage and talk about the language of mathematics. I leave the metalanguage unspecified in detail; to begin with I assume that it is rather weak, with just enough machinery to

Prof. Monk received his Berkeley Ph.D. in 1961 under Alfred Tarski. After a post-doctoral year at Berkeley, he came to his present post at Colorado. His main research is in algebraic logic, and he has published the books, *Introduction to Set Theory* (McGraw-Hill 1969) and (with L. Henkin and A. Tarski) *Cylindric Algebras, Part I* (North Holland, forthcoming). *Editor.*

formulate the above precise notions. If T is a list of sentences (thought of as axioms for a certain theory), and ϕ is a single sentence, I write $T\vdash\phi$ to indicate that there is a formal proof of ϕ from T, i.e., that there is a list of sentences, each of which is either a logical axiom, appears in the list T, or is obtained from earlier sentences of the list by applying a rule of inference. I write $T\nvdash\phi$ if there is no proof of ϕ from T. Before embarking on a description of set-theoretical matters I want to mention two famous theorems of Gödel which form a background upon which to view the latest results.

THEOREM 1. (Incompleteness theorem.) *If T is a consistent, sufficiently strong, effective list of sentences, then there is a sentence ϕ such that $T\nvdash\phi$ and $T\nvdash\neg\phi$.*

Here "$\neg$" is an abbreviation for "not." T is *consistent* if $T\nvdash\psi$ for some ψ. To say that T is *sufficiently strong* means, roughly speaking, that T embodies enough mathematics to develop elementary number theory; technically speaking, Peano arithmetic, P, is relatively interpretable in T. To say that T is *effective* means that the list T is presented in a reasonable manner—reasonable enough for oe tno be able to recognize by some algorithm when a sentence is in the list T. Certainly T is effective if T is finite: to check if a sentence ϕ is in T just look through the whole list, a process which in principle terminates with the last member of T. But Theorem 1 applies to some infinite lists also. Precisely speaking, one assigns numbers, called *Gödel numbers*, to all sentences, and T is called effective if the set of Gödel numbers of members of T is a recursive set. Clearly, to express this notion of effectiveness my original weak metalanguage must be strengthened enough to work in an elementary way with recursive functions, integers, and sets of integers. The other theorems of this section are also formulated in this stronger metalanguage, which is still much weaker than the ordinary language of mathematics.

Theorem 1 itself has a profound philosophical significance. According to it, one cannot hope to base all of conceivable mathematics on a single axiomatic basis; it points out the necessity of a continuous search for additional axioms. The importance of the theorem can be more appreciated in connection with the opposed philosophical views of formalism and platonism which will be discussed in section 2. I want to point out now, though, that the incompleteness theorem has not had much effect on the attitude of the working mathematician. In contrast, the latest results concerning the independence of the Axiom of Choice and the Continuum Hypothesis are already having an effect on teaching and research. I think that one of the main reasons for the lesser practical significance of the incompleteness theorem lies in the nature of the known proofs of the theorem. The sentence ϕ whose existence is asserted in the theorem *is* effectively constructed, but its intuitive meaning is not to be found in ordinary mathematics. It can be interpreted as asserting a relationship between formal proofs and its own Gödel number, and its construction is another instance of the well-known Cantor diagonal method. Theorem 1, essentially due to Gödel, can be

found in Tarski, Mostowski, Robinson [10] (it follows immediately from Theorems I1, I7, I10, and I19 there).

If T is an effective list of sentences, one can effectively construct a number-theoretic sentence $ConT$ which expresses the statement that T is consistent (under a natural interpretation, via Gödel numbers).

THEOREM 2. (Gödel's second underivability theorem.) *If T is a consistent, sufficiently strong, effective list of sentences, then* $T \nvdash ConT$.

See Feferman [2] for a proof of this theorem; it is also pointed out in [2] that one must take some care in the construction of the sentence $ConT$. This second theorem of Gödel again puts a limitation on what one can do in the foundations of mathematics. If T is a list of sentences which provides axioms for a large portion of mathematics, then by Theorem 1 not all mathematics is encompassed by T. One is thus forced to be somewhat modest. Theorem 2 forces a more severe degree of modesty: it is impossible to prove the consistency of T (without using devices not available in T itself). One would like to know that working with T has some significance, which it does not if T is inconsistent; so no method is available for logically proving that mathematics is significant.

I now turn to set theory. Students are taught these days that all mathematics can be based on set theory, indeed, that ordinary mathematics logically speaking is just a branch of set theory. It is, in fact, well established that almost all mathematics can be reduced to set theory. The only doubts that may arise concerning such a reduction have to do with the recently developed theory of categories. Anyway, in talking about set theory I think one is talking about essentially all of mathematics. I shall fix upon a particular list ZF of axioms for set theory, the *Zermelo-Fraenkel axioms*. These are essentially the axioms used in Bourbaki [1] and Halmos [5], but I assume that the axiom of regularity is also included (see Monk [8]), while the Axiom of Choice is excluded. By the *Axiom of Choice*, for short AC, I mean the statement that if A is a family of nonempty sets, then there is a function f (called a *choice function* for A) such that $f(x) \in x$ for each $x \in A$. If A is a finite family (*finite* defined conventionally), then such a function can be proved to exist within ZF, using only the simplest set-theoretical axioms (this was overlooked in Hall, Spencer [4], p. 282). In general, however, the existence of f requires infinitely many choices, and there is no principle within ZF which makes this legal. The axiom of choice can be proved equivalent, within ZF, to Zorn's lemma, to the Well-ordering Principle, to Tukey's lemma, etc. By the *Continuum Hypothesis*, for short CH, I mean the statement that any infinite set of real numbers can be placed in one-to-one correspondence either with the integers or with the set of all real numbers. Finally, the *Generalized Continuum Hypothesis*, GCH for short, asserts that for any infinite set X and any family Y of subsets of X, Y can be placed in one-to-one correspondence either with a subset of X or with the set of all subsets of X. Since in $ZF+AC$ one can show that the reals can be placed in one-to-one correspondence with the set of all subsets of Z (the set of integers), it follows that

$ZF+AC\vdash GCH\rightarrow CH$. Now I shall list a few of the important recent results in the foundations of set theory.

THEOREM 3. (Gödel, 1938.) $Con(ZF)\Rightarrow Con(ZF+AC+GCH)$.

I could just as well formulate Theorem 3 as follows:

If ZF is consistent, then so is $ZF+AC+GCH$.

Theorem 3 is written as a theorem of number theory as so to make more precise the assumptions in the metalanguage and to establish a connection with Theorem 2. The question of consistency of $ZF+AC+GCH$ reduces to the same question for ZF. By Theorem 2 this latter question cannot be given a rigorous affirmative answer. Since set theory, in particular ZF, has been used so much in the last century, mathematicians have grown confident that it is, in fact, consistent. Thus one can assert with the same degree of confidence that $ZF+AC+GCH$ is consistent.

THEOREM 4. (Cohen, 1963.) $Con(ZF)\Rightarrow Con(ZF+AC+\neg CH)$.

In Theorem 4, $\neg CH$ can be taken in various very specific forms, such as $2^{\aleph_0}=\aleph_2$, or $2^{\aleph_a}=\aleph_3$, or $2^{\aleph_0}=\aleph_{\omega+1}$. The most general results of this sort have been established by Easton and Solovay; see, e.g. [9].

THEOREM 5. (Fraenkel, Mostowski, Cohen, 1929–1963.) $Con(ZF)\Rightarrow Con(ZF+\neg AC)$.

Again, $\neg AC$ can be taken in many more definite forms, for example: there is a countable collection of unordered pairs without a choice function.

THEOREM 6. (Vitali, 1905.) $ZF+AC\vdash \exists x$ (*x is a set of real numbers, but x is not Lebesgue measurable*).

THEOREM 7. (Solovay, 1965.) $Con(ZF'+AC)\Rightarrow Con(ZF+\neg AC+\forall x$ (*if x is a set of real numbers, then x is Lebesgue measurable*)).

Here ZF' is obtained from ZF by replacing the usual axiom of infinity by a stronger one which asserts the existence of an uncountable strongly inaccessible cardinal. This stronger axiom is coming more and more to be an accepted part of set theory. For example, category theory appears to require this axiom, or even stronger axioms, to justify its methods. Intuitively, $Con(ZF'+AC)$ seems as plausible as $Con(ZF+AC)$; like $ZF+AC$ itself, the consequences of $ZF'+AC$ have been rather well worked-out and some confidence can be placed in its consistency. Since, however, $ZF'+AC\vdash Con(ZF+AC)$, the new axiom of infinity is much stronger than the old axiom (see Theorem 2).

THEOREM 8. (Solovay.) $Con(ZF)\Rightarrow Con(ZF+\neg AC+\exists x$ (*x is a set of real numbers and x is not Lebesgue measurable*)).

By Theorem 8, the Axiom of Choice is not equivalent to the existence of non-Lebesgue-measurable sets of real numbers.

These theorems I have listed represent a small sampling of results known in this area; they are sufficient to form a basis for the non-mathematical arguments in the next section. The interested reader can think of many familiar theorems whose proofs involve the Axiom of Choice, for example, and ask whether results similar to the above for Lebesgue measure hold. Thus the existence of Hamel bases, the Hahn-Banach extension theorem, the Boolean prime ideal theorem, the Banach-Tarski paradox, and the extendability of a partial order to a linear order are theorems which give rise to independence questions of this sort. Even assuming the Axiom of Choice, many statements of a set-theoretical nature have until recently been open. Examples are: the existence of a nontrivial measure on the set of all subsets of a set, and Souslin's hypothesis. Some of these many natural hypotheses have now been settled (in the sense of being shown independent), while others are still under attack. The intuitive remarks in the next section apply to all of these questions.

2. Meaning of the results. What do these results "do" for the ordinary mathematician? Before indicating some specific possibilities along these lines, it is worthwhile briefly to take a deeper view of the significance of the results. The results throw a great deal of light on a certain dichotomy in the philosophy of mathematics which now has a long history. Without trying to connect theories in the philosophy of mathematics with broader philosophical trends, I will distinguish two extreme views, *platonism* and *formalism.* These are not the only possible philosophies of mathematics. For example, intuitionism has a great appeal and is close to the beliefs of many practicing mathematicians. But the philosophies of most mathematicians can be construed as somewhere in the range between extreme platonism and extreme formalism. Practicing mathematicians, consciously or not, subscribe to some philosophy of mathematics (if unstudied, it is usually inconsistent). If you make a simple reference to the real numbers, you express a tendency toward platonism. And if you refer to a theorem as correct because it follows from the axioms of set theory, you tend toward formalism.

According to extreme platonism, mathematical objects are real, as real as the world we live in. Thus infinite sets exist, not just as a mental construct but in a real sense, perhaps in a "hyperworld." Similarly, nondenumerable sets, real numbers, choice functions, Lebesgue measure, etc., have a real existence. From the point of view of platonism, the purpose of a mathematician is to discover some of the facts of nature. His job is thus quite similar to that of a physicist, chemist, or biologist. The various possible axioms of set theory are then either true or false, and one of the main aims in the foundations of mathematics is to develop correct intuitions so as to determine which are the true axioms; these may then be taken as a rigorous basis for set theory. Actually, for a platonist, axiomatic development of set theory is not essential, but is perhaps useful to keep from making mistakes. (Even a platonist, however, will admit the importance of axiomatic treatments outside set theory, as in topology or group

theory, because of the usefulness of axiomatizations for abstractions and classifications.)

A strict formalist, on the other hand, does not believe that any mathematical objects have a real existence. For him, mathematics is just the business of deriving sentences from axioms. It is a game, in that some definite rules must be followed in such derivations. Unsolved problems give rise to goals for the game; the winner is the one who solves the problem. The analogy with games like chess and go is very close. A formalist chooses which game to play, that is, which axioms to take and which problem to work on, using practical and artistic criteria. One set of axioms may be best suited to be a base for a physical theory like relativity, for example, and hence because of the predictive ability of the physical theory this set of axioms has a practical value. And, of course, some problems are more practical in nature than others. The criteria for choosing axioms and problems, when not practical in nature, are extremely varied. A certain axiom may enable one to resolve many questions that are difficult without its aid; assuming that *GCH*, for example, infinite cardinal arithmetic is very much simplified. Many investigations are made in order to try to relate two seemingly distant areas in mathematics; one may cite the duality theory for Boolean algebras, relating algebraic to topological structures, or the investigation of closure algebras—doing topology within algebra. These two criteria for choosing the right game are just examples, and are undoubtedly not the most important of those criteria which are not based on practical considerations.

The results of section 1 have different meanings for platonists and formalists. The incompleteness theorem (Theorem 1) shows the platonist that he cannot hope to capture all of mathematics in a completely rigorous form for once and for all. There will remain beyond any fixed rigorous framework a statement whose truth must be determined by intuition. On the other hand, a strict formalist may even doubt that Theorem 1 says something relevant to his activity, since, as I indicated in section 1, the formulation and proof of Theorem 1 require a metalanguage stronger than the minimal one needed to understand the notions of proof and theorem which are the basic "rules of the game" for the formalist. But if the formalist does admit the usual intuitive meaning of Theorem 1, the theorem will just be taken as evidence that one cannot be content with just one axiom system if one wants to develop a comprehensive part of mathematics. To a platonist, Theorem 2 shows again the weakness of axiom systems; to him a system such as ZF, for example, is obviously consistent, since all the axioms of ZF are intuitively true. A formalist would also view Theorem 2 as showing a weakness of formal systems, again, if he admitted the usual intuitive meaning of the theorem. The other results in Section 1, independence results in set theory, give several examples of important statements which cannot be decided on the basis of the usual axioms of set theory. Here the platonist will try to investigate the situation further, in hopes of finding an impelling intuitive principle with the aid of which these statements can be resolved one way or the other. The formalist will simply view the results as giving

rise to several alternative paths of development of set theory.

These two philosophical viewpoints do not make much difference in the communication of mathematical results to other mathematicians. A correct mathematical argument from given premises is recognized as such by a platonist and by a formalist. The philosophical questions inherent in the dichotomy can be, and are, ignored in the writing of most mathematicians. In the main this indifference to the philosophy of mathematics has had a good effect on the progress of mathematics; mathematicians have insisted on proving theorems rather than spending a bulk of their time with difficult and ultimately inclusive philosophical speculations. Nonetheless, the two views have an effect on the direction of mathematical research. For example, a platonist may convince himself that CH is false (cf. Gödel [3]). He is then less likely to try to derive consequences from CH. But a formalist may very well like CH and even GCH because he can prove many nice theorems with their aid.

My remarks so far in this section concern mathematics itself, or at least mathematical research. I now turn to some specific possibilities for change in teaching and research which might come about because of the recent independence results. Mainly I will discuss the definition of the real number system. In beginning analysis courses it is customary to give the main properties of the real numbers and perhaps to carry out one of the constructions of the reals from the rationals. Many constructions of the reals are known; the ones using Dedekind cuts and Cauchy sequences are the most popular. All of these constructions turn out to be equivalent. This fact is rightly used, I think, as evidence for the naturalness of the notion of real number. Another basic, and satisfying, result here is uniqueness: any two Dedekind-complete ordered fields are isomorphic. But what are the real numbers? Under a platonistic point of view, the real numbers exist in nature; in teaching beginning mathematicians it would be nice to be able to point and say: here they are. A natural definition would be as an isomorphism equivalence class of Dedekind complete ordered fields. However, such classes are too big, and are not admitted as existing in ZF. There is a sophisticated way of chopping such classes down to a manageable size, but the method is not suitable for elementary classes (cf. Monk [8], p. 114). The only natural way out seems to be to fix upon a definite construction of the reals in order to give a specific definition for them. Then different mathematicians will have different definitions, but at least they can be shown equivalent.

Frequently the uniqueness theorem is used as a basis for asserting that all questions about the real numbers can be resolved, at least theoretically: anything true of one Dedekind-complete ordered field is true of another. But then Theorems 3 and 4 pose a puzzle. CH is a property of the real numbers (in a broad sense), and it is consistent to assume CH, but also consistent to assume $\neg CH$ (assuming ZF consistent). A closer analysis reveals the true state of affairs. Call a Dedekind-complete ordered field *a system of real numbers.* It is then provable within ZF that CH holds with respect to one system of real numbers if and only if it holds with respect to any other system of real numbers.

But neither side of this biconditional is *actually* provable in ZF. One may say that the uniqueness theorem (theoretically) reduces all questions about the reals to purely foundational, set-theoretical questions. It seems to me to be appropriate to bring discussions such as this down to the teaching level. Students should be aware of the possibility of getting different conceptions of the real numbers by choosing one or another of various hypotheses such as CH. It also seems to me that it would be useful to make students aware of the alternative philosophies of platonism and formalism.

Similarly, in the important applications of the Axiom of Choice I think it is appropriate to point out various alternatives that exist. Above all, I think a retreat from dogmatism is called for. In proving the existence of non-Lebesgue-measurable sets, it should be pointed out that, at a price, one can as well assume that every set of real numbers is Lebesgue-measurable. Similar remarks are appropriate whenever the mathematical or foundational results are of significant import to most mathematicians; certainly when one discusses such topics as the Hahn-Banach theorem, Souslin's hypothesis, the existence of maximal ideals, Tychonoff's theorem, etc.

The possibilities for remarks in classroom teaching thus appear to be very great, even though one cannot expect of teachers more than a mention of the independence results in question, since the proofs seem as yet inaccessible to an audience not able to devote some months to a study of these matters.

Possible uses of these results in research are rather obvious, but limited. Thus most mathematicians do not work in areas where a choice of CH or $\neg CH$ would make a big difference. And most modern mathematics depends fundamentally on the Axiom of Choice, so that the independence results such as I have mentioned are not of great practical import. But there are some instances in research where a new foundational hypothesis might prove useful. An analyst might like to assume that every set of real numbers is Lebesgue measurable. This can be done even while retaining a weak form of the Axiom of Choice. Again, the assumption that $2^{\aleph_0} = \aleph_2$ might facilitate the construction of counterexamples in certain contexts. Research based on Souslin's hypothesis has not been done very much. Many more possibilities could be stated, and research using these unusual hypotheses is needed. From a platonistic point of view such research might lead to a better insight into the nature of our underlying set theory.

I hope that the discussion I have given will convince more mathematicians to become familiar with the results obtained in the foundations of set theory and keep these results in mind in their teaching and research.

This article is the approximate text of a talk given at the Rocky Mountain Section meeting of the Mathematical Association in May, 1969. I wish to thank Gebhard Fuhrken and Watson Fulks for useful comments on a first draft of this article.

References

1. N. Bourbaki, Théorie des Ensembles, Chapters I and II, Hermann, Paris, 1954, 137 pp.

2. S. Feferman, Arithmetization of metamathematics in a general setting, Fund. Math., 49 (1960) 35–92.

3. K. Gödel, What is Cantor's continuum problem? this MONTHLY, 54 (1947) 515–525.

4. D. W. Hall, and G. L. Spencer, II, Elementary Topology, Wiley, New York, 1955, xii+303 pp.

5. P. R. Halmos, Naive Set Theory, Van Nostrand, Princeton, N. J., 1960, vii+104 pp.

6. I. Lakatos, editor. Problems in the Philosophy of Mathematics, North-Holland, Amsterdam, 1967, xv+241 pp.

7. A. R. D. Mathias, A survey of recent results in set theory, Amer. Math. Soc. Proc. Pure Math., 13 (1970) Axiomatic Set Theory, to appear.

8. J. D. Monk, Introduction to Set Theory, McGraw-Hill, New York, 1969, ix+193 pp.

9. R. Solovay, $2^{\aleph_0}$ can be anything it ought to be, Theory of Models, North-Holland, Amsterdam, 1964, p. 435.

10. A. Tarski, A. Mostowski, and R. M. Robinson, Undecidable Theories, North-Holland, Amsterdam, 1953, xi+98 pp.

THINKING GEOMETRICALLY*

DANIEL PEDOE, University of Minnesota

To many people the word "geometry" inevitably suggests a figure, a drawing. We are aware of the fact, apparently overlooked by Euclid, that we have to be very careful in arguing from a figure, that we may unwittingly assume properties which are not deducible from the given hypotheses, and may therefore arrive at incorrect logical conclusions. This may be a partial explanation of the fact that the whole subject of geometry, especially elementary geometry, is under attack these days. The leader of the attack, and a very formidable person he is, seems to be my old friend Prof. Jean Dieudonné. He is, of course, a very fine geometer, and a well-known member of the Bourbaki school.

Dieudonné has made his views known on a number of occasions, and most explicitly perhaps in a long preface to a book *Linear algebra and geometry*, pub-

* This paper is based on lectures given at the Universities of South Carolina and Toronto in the spring of 1968, and at Makerere College, Uganda and the University of the Witwatersrand in the summer of 1969. The invited address to the North Central Section of the MAA in April, 1969 contained similar material.

Prof. Pedoe studied at the Universities of London, Cambridge, and the Institute for Advanced Study. He held instructorships at Southampton, Birmingham, and London, a readership in the Univ. of London, and Professorships at Khartoum, Singapore, Purdue University, and his present post, the Univ. of Minnesota.

Dan Pedoe is well known for his elegant geometrical expositions in many articles, films, and books. The latter include the 3 vol. *Methods of Algebraic Geometry* (with W. Hodge, Cambridge U. Press, 1947–1953), *Circles* (Pergamon Press, 1957), *Gentle Art of Mathematics* (English U. Press, 1958, Penguin, 1963), *Geometric Introduction to Linear Algebra* (Wiley, 1963), *Introduction to Projective Geometry* (Pergamon, 1964), and *Course of Geometry for Colleges and Universities* (Cambridge U. Press, forthcoming). He received an MAA Lester Ford Award in 1968. *Editor.*

lished in 1964, and recently reissued in a third edition and in an English translation [1]. This book was written as a teachers' book for high schools in France, and introduces geometric ideas in the plane and Euclidean space via linear algebra. There are no diagrams, although their use is not forbidden in this book as they are in a more advanced one, where Dieudonné talks of a "strict adherence to axiomatic methods, with no appeal whatsoever to 'geometric intuition', at least in the formal proofs: a necessity which we have emphasized by deliberately abstaining from introducing diagrams in this book. My opinion is that the graduate student of today must, as soon as possible, get a thorough training in this abstract and axiomatic way of thinking," [2, p. 5].

The high school book begins with a set of axioms for the real numbers, then with a set for vector spaces, and proceeds to develop linear algebra and to introduce the geometry of the Euclidean plane and space by these methods. Dieudonné says that he knows nothing about the way children between the ages of 11 and 14 think, but he thinks that the Hilbert axioms by which Euclidean geometry can be developed rigorously are too involved, and that a development via a simpler set of axioms is desirable. His book is written, he adds, merely to put on record for the benefit of any future historian who might be interested how Dieudonné believes that elementary geometry can be taught in a rational manner. It should be said, parenthetically, that using this book to teach from is a most stimulating intellectual adventure.

In case the reader might not notice certain omissions in his book, Dieudonné makes it plain in his preface that he thinks French schools spend far too much time on special properties of the triangle, on trigonometry, on circles and systems of circles, on conics and systems of conics, and so on. He is very amusing when he says that books on trigonometry filled with formulas are written for astronomers, surveyors, and for writers of books on trigonometry, and that school children should never be deliberately trained for any of these professions! He gives, in his book, the essentials of trigonometry for use in modern mathematics, and there are no special formulas for the triangle. In fact, at a meeting of the North Atlantic Treaty Organization held some years ago, Dieudonné went on record as saying "Away with the triangle!" He does not want triangles mentioned, let alone studied in elementary geometry at all.

On another occasion Dieudonné has asked: "Who ever uses barycentric coordinates?", and in the preface to the high school book he complains that he studied systems of circles as a student, but has never come across them again in mathematics. If, says Dieudonné, one ever comes across a conic, one can treat it by the methods of the differential and integral calculus, as one does other curves.

In the course of this paper I shall discuss some problems—and mathematics, after all, is concerned with posing and solving problems—in which ideas about triangles, barycentric coordinates, systems of circles and conics help in the solution. I shall not use the differential calculus.

Perhaps I should assure any anxious reader that there is no evidence that

too much time is spent in American high schools on the study of triangles, circles, and whatnot. I asked my sophomore students recently, working backwards to try to discover something they did know, whether anyone knew the formula for the area of a triangle. Eventually one student said: "One half the base times the height," at which another student remarked very wisely: "Oh, I only had that for a right-angled triangle!"

1. The first problem I wish to discuss in which geometrical thinking is helpful runs as follows: Perpendiculars are dropped from the vertices A, B, C of a triangle ABC in a plane π onto a distinct plane π', forming an equilateral triangle $A'B'C'$. Show that the length of a side of $A'B'C'$ satisfies the equation

$$3x^4 - 2x^2(a^2 + b^2 + c^2) + 16\Delta^2 = 0,$$

where a, b, c are the lengths of the sides of the given triangle ABC, and Δ is its area.

This is a straightforward exercise in the use of Pythagoras. It occurred to me to wonder whether this equation has real roots. This was not mentioned in the problem, which turned up as an exercise in an English high school textbook. The condition is $a^2+b^2+c^2 \geqq 4\sqrt{3}\Delta$; with the help of a little algebra, this condition is equivalent to the condition

$$(b^2 - c^2)^2 + (c^2 - a^2)^2 + (a^2 - b^2)^2 \geqq 0.$$

This shows that the roots of the equation are always real, and that the inequality for a triangle, which is therefore always valid, becomes an equality if and only if the triangle ABC is equilateral.

This did not seem to be terribly exciting, until I discovered that in 1919 Weitzenboeck [**10**] had published this inequality for a triangle, had proved it algebraically, and was obviously rather proud of his result. It had occurred to me in the meantime that a better way to show that the roots of the equation are always real is to inscribe an ellipse touching the sides of the triangle ABC at their respective midpoints, and to project this ellipse orthogonally into a circle. This is always possible, and the resulting triangle $A'B'C'$ around the circle must be an equilateral triangle. This projection is the basis of a film I have made called *Orthogonal Projection*, and the problem of inscribing an ellipse in a triangle to touch the sides at their midpoints is the basis of another film called *Central Similarities* [**9**].

Having arrived at an inequality for any triangle by showing that an orthogonal projection into an equilateral triangle is always possible, I wondered whether a given triangle ABC can always be projected orthogonally into a triangle of *given shape*, with sides ka', kb', kc'. The method, using Pythagoras, is as before and one ends with the equation

$$16\Delta'^2k^4 - 2\Theta k^2 + 16\Delta^2 = 0,$$

where Δ is the area of triangle ABC, where Δ' is the area of a triangle with sides

a', b', and c', and where

$$\Theta = a^2(-a'^2 + b'^2 + c'^2) + b^2(a'^2 - b'^2 + c'^2) + c^2(a'^2 + b'^2 - c'^2).$$

The condition for real roots of this equation is

$$\Theta^2 \geqq (16\Delta\Delta')^2,$$

and this suggested the possibility of a two-triangle inequality:

$$\Theta \geqq 16\Delta\Delta'.$$

I wondered whether I could show geometrically that the orthogonal projection is always possible, and so derive the two-triangle inequality. This is where barycentric coordinates (areal coordinates) are useful!

Taking ABC as triangle of reference, consider the inscribed conic with equation

$$p^2X^2 + q^2Y^2 + r^2Z^2 - 2pqXY - 2qrYZ - 2rpZX = 0.$$

If we consider the intersections of this conic with $X+Y+Z=0$, the line at infinity, the discriminant of the resulting quadratic equation is

$$-pqr(p + q + r),$$

and so we are assured that if p, q, r are all positive, that is, if the points of contact of the conic with the sides of ABC are all internal points, then the inscribed conic is an ellipse. Project this ellipse into a circle by orthogonal projection, and suppose that the triangle ABC projects into triangle $A'B'C'$. The center of the ellipse projects into the center of the incircle for triangle $A'B'C'$. The center of the ellipse has coordinates

$$q + r : r + p : p + q,$$

and the center of the circle has coordinates $a':b':c'$, where these are the lengths of the sides of triangle $A'B'C'$. Since the ratio of areas is unchanged by orthogonal projection, and the center of the ellipse projects into the center of the incircle, we have

$$q + r : r + p : p + q = a':b':c',$$

so that

$$p : q : r = -a' + b' + c' : a' - b' + c' : a' + b' - c'.$$

If we choose the ratios $p:q:r$ to satisfy these equations and project the resulting ellipse orthogonally into a circle, then the projection of triangle ABC will be a triangle similar to triangle $A'B'C'$.

We have not yet finished with the geometrical ideas associated with this two-triangle inequality. In 1937 Finsler and Hadwiger pursued the ideas of Weitzenboeck, and derived his triangle inequality as follows: [3]. They erected equilateral triangles $A''BC$, $B''CA$ and $C''AB$ inwards on the sides BC, CA

and AB respectively of triangle ABC. Let V be the centroid of $B''CA$ and W the centroid of $C''AB$. Then by trigonometry,

$$(VW)^2 = k(a^2 + b^2 + c^2 - 4\sqrt{3}\Delta),$$

where k is a positive constant for the triangle ABC. Since $(VW)^2 \geqq 0$, we have $a^2+b^2+c^2 \geqq 4\sqrt{3}\Delta$, with equality if and only if the triangle ABC is equilateral.

I was certain that this method would give my two-triangle inequality. I erected triangles $A''BC$, $B''CA$, and $C''AB$ inwards on the sides BC, CA and AB respectively of triangle ABC, each similar to a triangle $A'B'C'$, and then I had to determine what would correspond to the centroid of an equilateral triangle for these triangles $A''BC$, $B''CA$ and $C''AB$. Anyone who has worked with triangles knows that the centroid has no useful angles associated with it, but the circumcenter of a triangle has, and I found that the square of the distance between the circumcenter V of $B''CA$ and the circumcenter W of $C''AB$ is given by the formula

$$2(VW/a')^2 = (R'/a'b'c')^2\left[\sum a^2(-a'^2 + b'^2 + c'^2) - 16\Delta\Delta'\right],$$

where R' is the circumradius of triangle $A'B'C'$. This shows that the triangle UVW is similar to the triangle $A'B'C'$, and also gives the two-triangle inequality

$$a^2(-a'^2 + b'^2 + c'^2) + b^2(a'^2 - b'^2 + c'^2) + c^2(a'^2 + b'^2 - c'^2) \geqq 16\Delta\Delta',$$

with equality if and only if the two triangles ABC, $A'B'C'$ are similar. This further condition comes immediately from the construction. I imagine that if Finsler and Hadwiger had thought of this, they would have been rather excited, since they spent so much time on proving the inequality for one triangle. If one of the two triangles is already equilateral, the two-triangle inequality simplifies to the triangle inequality for the other triangle. There is a much easier method for obtaining it [5], of course.

One final remark on this discovery. Since my note on the two-triangle inequality mentioned orthogonal projection, it was rejected by a journal I sent it to, as being "unsuitable." I rewrote the note, and gave the derivation from the method of Finsler and Hadwiger, illustrious names, after all, and obtained publication in a "Research Note" [6]. This was in 1943, and things published then did not travel, but I think that this discovery of the first interesting inequality for two triangles does show how geometrical thinking, necessarily based on some knowledge of geometry, can be fruitful.

2. My second look at geometrical thinking is bound up with what is called the Six-Circle Theorem. We have four circles C_1, C_2, C_3, C_4 in the inversive plane, and we suppose that C_1, C_2 intersect in the points P_1, Q_1, that C_2, C_3 intersect in P_2, Q_2, that C_3, C_4 intersect in P_3, Q_3, and that C_4, C_1 intersect in P_4, Q_4. The theorem says that if the four points P_1, P_2, P_3, P_4 lie on a circle, then so do the four points Q_1, Q_2, Q_3, Q_4. (Figure 1.)

This is a well-known theorem, and can be proved in a number of ways. It is

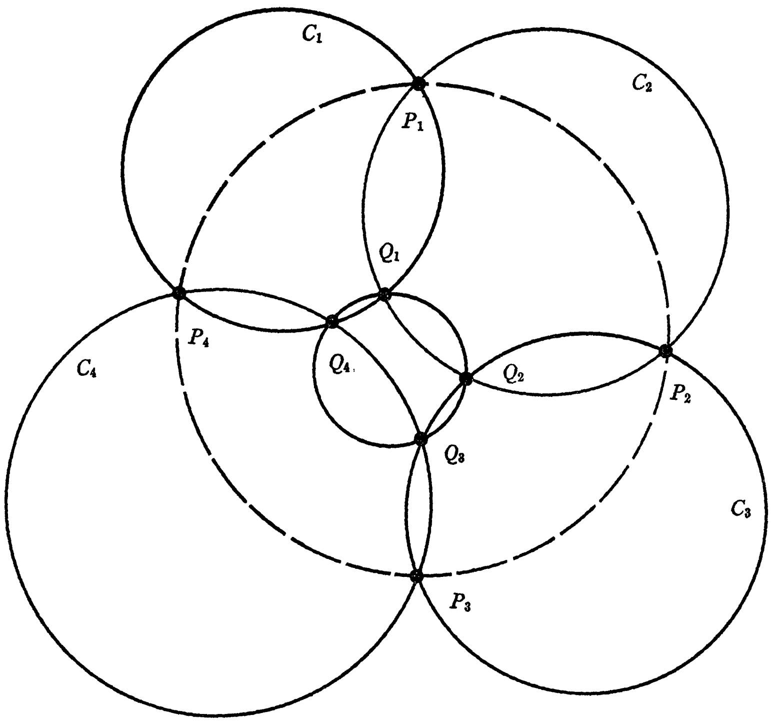


FIG. 1

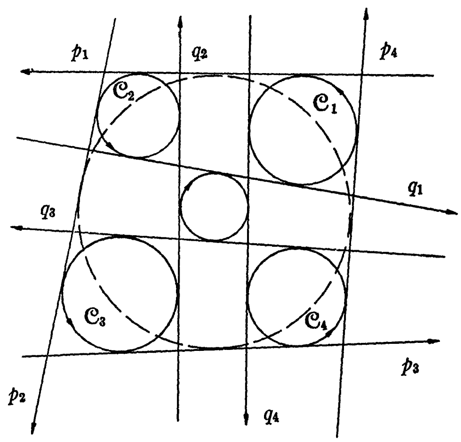


FIG. 2

an exciting theorem, and I am not just representing my own state of mind about it. It makes even the Russian writer Jaglom excited. He gives it in his recent rather curious book *Complex numbers in geometry* [4, p. 35], where he says:

"This proposition seems rather elegant, but not particularly promising—an ordinary theorem of which there are many in elementary geometry. However, the consequences which follow from this simple theorem can surely be called remarkable. As the first of these consequences we mention a whole series of theorems due to the English geometer Clifford." One chain of theorems begins with three lines, which determine three points by their intersections. There is a unique circle through these points. Four lines determine three sets of three lines, and therefore determine four circles. As a consequence of the Six-Circle Theorem these four circles meet in a point, which Jaglom calls the central point of the four lines. Five lines determine five sets of four lines, and each tetrad determines a central point. These five central points can be shown to lie on a circle, which Jaglom calls the central circle of the five lines, and so on. We are in the inversive plane, and must accept the fact that all lines contain the point at infinity, so that in the Six-Circle Theorem any one of the circles may be a line, and among the intersections of two lines there is always the point at infinity.

I agree with Jaglom that complex numbers afford a ready means of proving theorems in plane geometry, but in his book he goes on to develop the theory of dual complex numbers at great length. These are numbers of the form $a+\epsilon b$, where a, b are real and $\epsilon^2=0$. The application is to oriented circles, that is to circles with a sense, and he obtains, after many pages of work, another attractive theorem, on oriented circles and their tangents. If $\mathcal{C}_1$, $\mathcal{C}_2$, $\mathcal{C}_3$, $\mathcal{C}_4$ are four oriented circles, and the tangents to $\mathcal{C}_1$, $\mathcal{C}_2$ are p_1, q_1, the tangents to $\mathcal{C}_2$, $\mathcal{C}_3$ are p_2, q_2, the tangents to $\mathcal{C}_3$, $\mathcal{C}_4$ are p_3, q_3, and the tangents to $\mathcal{C}_4$, $\mathcal{C}_1$ are p_4, q_4, then if there is an oriented circle which touches the oriented lines p_1, p_2, p_3, p_4, there is also an oriented circle which touches the oriented lines q_1, q_2, q_3, q_4. (Fig. 2.)

I shall show that both the Six-Circle Theorem and the theorem on oriented circles arise quite naturally if we map the circles of the Euclidean plane onto the points of E_3. We map the circle with equation

$$X^2 + Y^2 - 2pX - 2qY + r = 0$$

onto the point (p, q, r) of E_3. This is a simple, yet fundamental mapping (it is equivalent to stereographic projection), and it does illustrate what Henri Lebesgue has referred to as one of the important aspects of geometry, that it is almost a plastic art. Seeing theorems from different aspects, I believe, is one of the great charms of geometry. In the mapping I have just described, those systems of circles called *coaxial systems*, or *pencils of circles* (given by the equation $kC+k'C'=0$, where C and C' are given circles, and k, k' vary over the real numbers) are mapped onto the lines of E_3. Is there anything more natural than a line in E_3? There appears a quadric Ω of equation $X^2+Y^2-Z=0$, and the points of this quadric represent the point-circles of the plane, those with zero radius. We soon find ourselves considering theorems which are obviously theorems of projective space, and there is a most attractive interplay between theorems for circles in the plane and theorems on points and lines in projective space, S_3. All this is worked out in [7], and in more detail in [8].

The Six-Circle Theorem becomes the following theorem in S_3. Suppose A, B, C, D are four distinct points in S_3, suppose Ω is a given quadric, and suppose AB intersects Ω in P_1, Q_1. Assume also that the line BC intersects Ω in the points P_2, Q_2, the line CD intersects Ω in the points P_3, Q_3, and finally the line DA intersects Ω in the points P_4, Q_4. Then if the points P_1, P_2, P_3, P_4 lie in a plane, so do the points Q_1, Q_2, Q_3, Q_4.

This theorem can be proved by using one of the most remarkable theorems on quadric surfaces in S_3, the theorem of the eight associated points. This says that if three quadrics intersect in eight distinct points, then any quadric which is made to pass through seven of these points will automatically pass through the last point of the eight. The eight points P_i, $Q_i (i=1, 2, 3, 4)$ are associated, since we can find three distinct quadrics which contain them. One is the quadric Ω itself, another is the reducible quadric which consists of the planes ABD and BCD, and a third is the reducible quadric which consists of the planes ABC and ADC. The two reducible quadrics intersect in the lines AB, BC, CD, and DA, so the three quadrics have only the eight points P_i, Q_i in common. Let the plane in which the four points P_i lie be denoted by π, and let the plane $Q_1Q_2Q_3$ be denoted by π'. Then the reducible quadric formed by the planes π and π' contains seven of the eight associated points, and must therefore contain the eighth point Q_4. If this does not lie in π, it must lie in π'. But it is seen immediately that if Q_4 is coplanar with the points P_1, P_2, P_3, P_4, then the points A, B, C, D are coplanar, and the theorem is trivial.

Of course, since we are using the theorem of the eight associated points in all its strength, we must be certain that our proof of this theorem holds for all possible cases. Unfortunately, few textbooks give an acceptable proof.

In S_3 we have a Principle of Duality, and so our theorem has an evident dual. When we write this down, and consider how oriented circles can be represented in E_3, this theorem for S_3, interpreted in E_3, gives us the theorem which Jaglom only proves after many pages of work on dual complex numbers. I shall not give the details here. They can be found in [8].

3. My last example on thinking geometrically is about conics, and I shall leave the question open as to whether the methods of the calculus would give a complete solution of the following problem. Given five distinct points in the Euclidean plane, we can draw a unique conic through them. If we begin with six distinct points, there are six conics which can be drawn to pass respectively through sets of five of the six given points. The problem is: can the six points be chosen so that (a) the six conics are all distinct ellipses, or (b) the six conics are all distinct hyperbolas, or (c) the six conics are all distinct parabolas? This question was asked some years ago by C. V. Durell, senior mathematics master at Winchester, one of Britain's great public (private) schools. The solution was given at sight by Dr. Beniamino Segre when I showed him the problem. Dr. Segre, who was then in England, is an illustrious successor of the great Italian geometers Corrado Segre, Guido Castelnuovo and Francesco Severi. The Segre

theorem is that it is possible to find $n>5$ points in the Euclidean plane so that the $\binom{n}{5}$ conics through sets of five of them are either all distinct ellipses or distinct hyperbolas, but it is impossible to find six points so that the conics through sets of five of them are all distinct and parabolas.

The proofs of the first two statements are similar, but the proof of the third part of the theorem has a completely different flavor, as we shall see. If we consider five points (x_i, y_i) $(i=1, 2, 3, 4, 5)$, we can obtain the equation of the conic which passes through them in various ways, by writing down a determinant, for instance. We are interested only in the highest degree terms. If these are $AX^2+2HXY+BY^2$. then the conic is an ellipse if and only if $H^2-AB<0$, and an hyperbola if and only if $H^2-AB>0$. In terms of the coordinates of the five points, we have the condition

$$\phi(x_1, y_1, \cdots, x_5, y_5) > 0,$$

where ϕ is a polynomial in the ten coordinates (x_i, y_i), as the condition for an ellipse, and $-\phi>0$ as the condition for an hyperbola. Since a polynomial is a continuous function of its variables, we know that we can draw small circles with their centers at the points (x_i, y_i) such that if $\phi>0$ at the points (x_i, y_i), then we still have $\phi>0$ at all points within the neighborhoods of the points. Now, the geometrical idea is to begin with all the n points on an ellipse. Then all the conics through sets of five of the points are ellipses, the same ellipse, and so $\phi>0$ for each set of five points. We now only have to vary the n points so as to keep $\phi>0$, but also so as to obtain $\binom{n}{5}$ distinct ellipses. The treatment for hyperbolas is similar; of course, the details have to be filled in. You may say that this is essentially a calculus proof, that is, one which uses analysis. The proof that we cannot find six points to produce six parabolas is quite different.

As soon as a geometer of what I must call the old school (I am one of them) sees six points in a plane, he thinks of the plane representation of a general cubic surface in S_3, in which plane sections of the surface are mapped onto cubic curves in the plane which pass through the six given points. Conics through five of the six given points intersect cubics through the six points in $2\cdot 3-5=1$ variable point, and since the cubic curves represent plane sections of the surface, the conics represent lines on the surface. The points themselves through which the cubics are drawn also represent lines on the surface, so we have twelve lines of what is called *a double-six* on the surface; two sets of six lines, each of which meets five lines of the other set. This information is not relevant to the problem, but merely indicates a fascinating branch of projective geometry. If the conics through sets of five of the six given points are all parabolas, then they all touch a special line in the plane. A line in the given plane meets the variable cubic curves through the six points in three variable points, and therefore represents a space cubic curve on the cubic surface. We therefore have the situation, if there are indeed six parabolas, that the cubic surface contains a space cubic curve which touches the six lines of one half of a double six of lines on the surface. We show that this assumption leads to a contradiction.

If we are skilled in the use of birational mappings of the projective plane (Cremona transformations), we can map the sextuplet of lines on the surface represented by the conics through sets of five points of the plane onto six points of the plane, and see what happens. The space cubic curve, originally mapped onto a line of the plane, becomes an irreducible curve of order five with cusps at the six fundamental points which now represent the parabolas through the original sets of five points. The order is five, because cubics through the six new points still represent plane sections of the surface, and if $3N-6\cdot 2=3$, we have $N=5$. The singularities must be cusps because the space cubic touches the lines which are now mapped onto points of the plane.

But there is no irreducible curve of order five with cusps at six distinct points. If we apply Pluecker's Formula to obtain the number of variable tangents which can be drawn to this quintic curve from an arbitrary point of the plane, we obtain the number $5\cdot 4-6\cdot 3=2$. It can be proved, however, that the only irreducible curve to which two variable tangents can be drawn from an arbitrary point in the plane is a conic. We therefore have a contradiction.

4. Not very long ago a mathematician, hot in the pursuit of modern mathematics, complained to me that it is not easy to find a book which deals with the plane representation of the cubic surface, and he needed it for his work. I met a young mathematician in Oxford recently who has studied the new algebraic geometry, but feels he must now learn some geometry. And there are others who have the new algebraic techniques at their finger tips but are searching for geometric ideas to motivate their research. There is no doubt that geometry should not be neglected, although the feeling against it in some quarters is almost Freudian in its intensity. The grudges arise from having been taught too much geometry, or having been taught geometry very badly. There is no reason why a course in geometry should not be well taught, and the danger of having too much is no longer a serious one. The United States is passing through a transition period in the teaching of geometry, and it should soon catch up with other countries, such as the Soviet Union, which have always given a lot of attention to the teaching of geometry. If one could find nothing else to say about geometry, it has always been conceded that students enjoy it, and the Pleasure Principle should not be neglected in mathematics. I have tried to communicate this feeling in my recent book [8], in which I have put down the geometric ideas which turn up in mathematics.

References

1. Dieudonné, Linear Algebra and Geometry, Hermann, Paris, 1964.
2. ———, Foundations of Modern Analysis, Academic Press, New York, 1960.
3. Finsler and Hadwiger, Comment. Math. Helv., 10 (1937) 316–26.
4. Jaglom, Complex Numbers in Geometry, Academic Press, New York, 1968.
5. Pedoe, Problem E 1562, this MONTHLY, 70 (1963) 1012.
6. ———, Proc. Cambridge Philos. Soc., 38 (1943) 397–98.
7. ———, Circles, Pergamon Press, London, 1957.

8. Daniel Pedoe, A Course of Geometry for Colleges and Universities, Cambridge University Press, England (in the press).
9. University of Minnesota College Geometry Project.
10. Weitzenboeck, Math. Zeit., 5 (1919) 137–146.

THE WILLIAM LOWELL PUTNAM MATHEMATICAL COMPETITION

J. H. McKAY, Oakland University

The following results of the thirtieth William Lowell Putnam Mathematical Competition held on December 6, 1969 have been determined in accordance with the regulations governing the Competition. This competition is supported by the William Lowell Putnam Intercollegiate Memorial Fund left by Mrs. Putnam in memory of her husband and is held under the auspices of the Mathematical Association of America.

The first prize, five hundred dollars, is awarded to the Department of Mathematics of **Massachusetts Institute of Technology,** Cambridge, Massachusetts. The members of the team were Don Coppersmith, John J. Keary, and Jeffrey C. Lagarias; to each of these a prize of one hundred dollars is awarded.

The second prize, four hundred dollars, is awarded to the Department of Mathematics of **Rice University,** Houston, Texas. The members of the team were Alan R. Beale, David A. Cox, and James B. Hobelman; to each of these a prize of seventy-five dollars is awarded.

The third prize, three hundred dollars, is awarded to the Department of Mathematics of the **University of Chicago,** Chicago, Illinois. The members of the team were Robert B. Israel, David S. Fried, and Kiyoshi Igusa; to each of these a prize of fifty dollars is awarded.

The fourth prize, two hundred dollars, is awarded to the Department of Mathematics of **Harvard University,** Cambridge, Massachusetts. The members of the team were George Sicherman, Gerald I. Myerson, and Mark A. Mostow; to each of these a prize of fifty dollars is awarded.

The fifth prize, one hundred dollars, is awarded to the Department of Mathematics of **Yale University,** New Haven, Connecticut. The members of the team were Gregg J. Zuckerman, J. Lance W. Jayne, and Frederic B. Weissler; to each of these a prize of fifty dollars is awarded.

The five persons ranking highest in the examination, named in alphabetical order, are **Alan R. Beale,** Rice University; **Don Coppersmith,** Massachusetts Institute of Technology; **Gerald A. Edgar,** University of California at Santa Barbara; **Robert A. Oliver,** University of Chicago; **Steven Winkler,** Massachusetts Institute of Technology. Each of these has been designated as Putnam Fellows by the Mathematical Association of America and is awarded a prize of two hundred and fifty dollars.

The five persons ranking second highest in the examination, named **in**

alphabetical order, are *Avner D. Ash*, Harvard University; *Martin J. Cohen*, University of California at Los Angeles; *Alan S. Lederman*, California Institute of Technology; *Paul S. Viita*, Harvard University; and *Gregg J. Zuckerman*, Yale University. To each of these a prize of one hundred dollars is awarded.

The following teams, named in alphabetical order, won honorable mention: *University of Alberta*, the members of the team were Stephen Whitney, Eleanor McDonald, and John Mallet-Paret; *University of California at Los Angeles*, the members of the team were Martin J. Cohen, Paul F. Klembeck, and Michael D. Miller; *California Institute of Technology*, the members of the team were Thomas R. Davis, Jerry M. Feinberg, and Randolph S. Tuler; *Princeton University*, the members of the team were Allen H. Back, Peter Malcolmson, and Steven H. Weintraub; *Reed College*, the members of the team were Thor O. Wenzel, R. Neil Vance, and Joe P. Buhler.

Honorable mention is given to the following twenty-six individuals, named in alphabetical order: Allen H. Back, *Princeton University*; Allen E. Barnes, *Drexel Institute of Technology*; Kenneth A. Brakke, *University of Nebraska*; Joe P. Buhler, *Reed College*; David M. Christie, *Massachusetts Institute of Technology*; Thomas R. Davis, *California Institute of Technology*; Zbigniew Fiedorowicz, *Illinois Institute of Technology*; Douglas A. Hensley, *University of Kansas*; William L. Hibbard, *University of Wisconsin at Madison*; Dean Hickerson, *University of California at Davis*; Kiyoshi Igusa, *University of Chicago*; Robert Israel, *University of Chicago*; Robert J. Kimble, *United States Naval Academy*; Nicholas Littlestone, *Harvard University*; Haynes R. Miller, *Harvard University*; Gerald Myerson, *Harvard University*; James R. Paulson, *Princeton University*; Adbeel N. Quiñones, *New York University at Washington Square*; William H. Rowan, *Stanford University*; George Sicherman, *Harvard University*; Walter Rees Stromquist, *University of Kansas*; Jon Sussman, *San Diego State College*; Robert E. Tax, *University of Chicago*; Steven H. Weintraub, *Princeton University*; Steven Zucker, *Brown University*; Michael Zuker, *McGill University*.

The other individuals who were ranked in the top one hundred, arranged by college, are: John J. Mallet-Paret and Stephen L. Whitney, *University of Alberta*; Charles W. Kaufman, *Bates College*; Philip Trauber, *Brooklyn College*; Douglas M. Lublin, *Brown University*; Daniel R. Farkas, *University of California, Santa Barbara*; Peter Van Wyck Loomis, *University of California, Davis*; Michael D. Miller, *University of California, Los Angeles*; Jon Gregor Bjornstad and Steven G. Krantz, *University of California, Santa Cruz*; Jerry M. Feinberg and David J. Smith, *California Institute of Technology*; James Richard Spriggs, *Case Western Reserve University*; David S. Fried, David J. Saltman, and Frank Anthony Wilczek, *University of Chicago*; Martin H. Ellis, *City College of New York*; Robert B. Lumbert, *Dartmouth College*; Daniel E. Frohardt, *Grinnell College*; Ira M. Gessel, Mark A. Mostow, John P. Robertson, Jonathan M. Rosenberg, and Paul J. Weiner, *Harvard University*; James L. Hlavka, *Harvey Mudd College*; James J. Callahan, *Manhattan College*; Charles E. Blain, Scott S.

Brown, Harold G. Donnelly, Daniel C. Galehouse, John J. Keary, Jeffrey C. Lagarias, Wralf T. Penquin, and Kenneth P. Rietz, *Massachusetts Institute of Technology*; Douglas A. Leonard and Robert L. Scott, *University of Michigan*; Serge Hamelin, *University of Montreal*; Christopher A. Gurwood, *New York University, University Heights*; Joseph L. Dunn, *New York University, Washington Square*; Curtis D. Herink, *North Central College*; Kim B. Bruce and Everett L. Bull, Jr., *Pomona College*; Alan M. Weinstein, *Princeton University*; Erik R. Lockeberg, *Queens University*; R. Neil Vance, *Reed College*; David A. Cox and James B. Hobelman, *Rice University*; Jerrold W. Grossman, *Stanford University*; Benjamin J. Kuipers, *Swarthmore College*; Mark E. Kaminsky, *The Cooper Union*; Daniel A. Gautreau, Daryl N. Geller, James M. Kavanagh, Hugh D. Miller and Joseph G. Sunday, *University of Toronto*; George Schick Lueker, *Valparaiso University*; Kent Brothers, *University of Victoria*; David G. Potter, *University of Waterloo*; Paul F. Daniels, *Williams College*; Ronald R. Pannatoni, *University of Wisconsin, Madison*; Kenneth P. Baclawski, *University of Wisconsin, Milwaukee*; Frederick J. Bruch, *University of Wisconsin, Parkside*; Jockum L. Aniansson, Dale H. Peterson and Frederic B. Weissler, *Yale University.*

One thousand five hundred and one students from two hundred eighty-eight colleges and universities participated in the examination on December 6, 1969.

A listing of the top five hundred contestants may be obtained from the Director. The list, which includes addresses and expected dates of graduation, may be helpful to departments of mathematics in selecting graduate students.

The Questions Committee, consisting of Leo Moser (chairman), Albert Wilansky, and Warren Loud, prepared the problems (listed below) for the competition.

PROBLEMS. PART A

A-1. Let $f(x, y)$ be a polynomial with real coefficients in the real variables x and y defined over the entire x-y plane. What are the possibilities for the range of $f(x, y)$?

A-2. Let D_n be the determinant of order n of which the element in the ith row and the jth column is the absolute value of the difference of i and j. Show that D_n is equal to

$$(-1)^{n-1}(n-1)2^{n-2}.$$

A-3. Let P be a non-self-intersecting closed polygon with n sides. Let its vertices be $P_1, P_2, \cdots, P_n$. Let m other points, $Q_1, Q_2, \cdots, Q_m$ interior to P be given. Let the figure be triangulated. This means that certain pairs of the $(n+m)$ points $P_1, \cdots, Q_m$ are connected by line segments such that (i) the resulting figure consists exclusively of a set T of triangles, (ii) if two different triangles in T have more than a vertex in common then they have exactly a side in common, and (iii) the set of vertices of the triangles in T is precisely the set of $(n+m)$ points $P_1, \cdots, Q_m$. How many triangles in T?

A-4. Show that

$$\int_0^1 x^x dx = \sum_{n=1}^{\infty} (-1)^{n+1} n^{-n}.$$

(The integrand is taken to be 1 at $x=0$.)

A-5. Let $u(t)$ be a continuous function in the system of differential equations

$$\frac{dx}{dt} = -2y + u(t), \qquad \frac{dy}{dt} = -2x + u(t).$$

Show that, regardless of the choice of $u(t)$, the solution of the system which satisfies $x=x_0$, $y=y_0$ at $t=0$ will never pass through (0, 0) unless $x_0=y_0$. When $x_0=y_0$, show that, for any positive value t_0 of t, it is possible to choose $u(t)$ so the solution is at (0, 0) when $t=t_0$.

A-6. Let a sequence $\{x_n\}$ be given, and let $y_n=x_{n-1}+2x_n$, $n=2, 3, 4, \cdots$. Suppose that the sequence $\{y_n\}$ converges. Prove that the sequence $\{x_n\}$ also converges.

PART B

B-1. Let n be a positive integer such that $n+1$ is divisible by 24. Prove that the sum of all the divisors of n is divisible by 24.

B-2. Show that a finite group can not be the union of two of its proper subgroups. Does the statement remain true if "two" is replaced by "three"?

B-3. The terms of a sequence T_n satisfy

$$T_nT_{n+1} = n \qquad (n = 1, 2, 3, \cdots) \text{ and } \lim_{n\to\infty} \frac{T_n}{T_{n+1}} = 1.$$

Show that $\pi T_1^2=2$.

B-4. Show that any curve of unit length can be covered by a closed rectangle of area 1/4.

B-5. Let $a_1<a_2<a_3<\cdots$ be an increasing sequence of positive integers. Let the series

$$\sum_{n=1}^{\infty} 1/a_n$$

be convergent. For any number x, let $k(x)$ be the number of the a_n's which do not exceed x. Show that $\lim_{x\to\infty} k(x)/x=0$.

B-6. Let A and B be matrices of size 3×2 and 2×3 respectively. Suppose that their product in the order AB is given by

$$AB = \begin{bmatrix} 8 & 2 & -2 \\ 2 & 5 & 4 \\ -2 & 4 & 5 \end{bmatrix}.$$

Show that the product BA is given by

$$BA = \begin{bmatrix} 9 & 0 \\ 0 & 9 \end{bmatrix}.$$

SOLUTIONS PART A

The number in parentheses, immediately following the problem number, is the number of participants who received a score of 8, 9 or 10 (10 is maximum possible) on the problem. In the case of A-1, A-2, B-1 and B-2, this applies to all 1501 participants. For the other problems, the count applies only to the 839 qualifiers.

A-1 (15). The continuity of $f(x, y)$ implies that the range is connected (i.e., if a, b are in the range and $a<c<b$ then c is in the range). If the range is bounded above and below, then the polynomial $f(x, kx)$ is a constant for each value of k and thus $f(x, y)$ is the constant $f(0, 0)$. Thus the only possibilities are: (i) a single point; (ii) a semi-infinite interval with end-point; (iii) a semi-infinite interval without end-point; and (iv) all real numbers.

Examples are easily given for (i), (ii) and (iv). An example for (iii) is harder to find. One way is to have each cross-section of the surface (for fixed y) be a parabola with a minimum which decreases asymptotically toward some constant as y approaches $\pm\infty$. A suitable example is $(xy-1)^2+x^2$.

Comment: When the examination was printed it was believed that (iii) was not possible. The arguments used were erroneous. This extra complexity is the primary reason for the few high scores.

A-2 (286). Subtract the first column from every other column. Then add the first row to every other row. The last row now has all zeros except for $(n-1)$ in the first column. D_n is $(-1)^{n-1}(n-1)$ times the minor formed by deleting the first column and last row from the transformed determinant. This minor has only zeros below the main diagonal and thus is equal to the product of its diagonal elements. Hence the minor has value 2^{n-2} and $D_n=(-1)^{n-1}(n-1)2^{n-2}$.

Alternate Solution: From the bottom row of D_{n+1}, subtract $1/(n-1)$ times the first row and $n/(n-1)$ times the nth row. This shows that $D_{n+1} = -[2n/(n-1)]D_n$, for $n>1$. The result follows easily by iteration and the observation that $D_2=-1$.

A-3 (39). Let t be the number of triangles. The sum of all the angles is πt (since it is π for each triangle) and it is also $2\pi m+(n-2)\pi$.

Alternate Solution: Let t be the number of triangles. In Euler's formula $V-E+F=2$, $F=t+1$, and $V=n+m$. Since every edge is on two faces, $2E=3t+n$. Substitution leads directly to the answer $t=2m+n-2$.

Comment: It should have been stated in the problem that the *interior* of the polygon is triangulated. If any of the additional line segments are outside of the polygon, the answer is different.

A-4 (30). A reasonable way to get a series (other than using Riemann sums, which apparently doesn't work) is to write the integrand as a power of e and use the series expansion for e. Then uniform convergence can be applied to interchange integration and summation, and show that

$$\int_0^1 x^x dx = \sum_{m=0}^{\infty} \frac{1}{m!}\int_0^1 x^m(\log x)^m dx.$$

Let $F(m, k)=\int_0^1 x^m(\log x)^k dx$. Integration by parts shows, if applied to a typical term in the summation, why we are interested in $F(m, k)$ and also shows that $F(m, k)=-k/(m+1)F(m, k-1)$ for $m\geqq 0$ and $k\geqq 1$. As a result, $F(m, m) =(-1)^m m!(m+1)^{-m}F(m, 0)=(-1)^m m!(m+1)^{-m-1}$. To get the given formula in the problem, replace $m+1$ by n and adjust the limits on the summation accordingly.

A-5 (45). Subtracting the two equations eliminates $u(t)$ and provides the

simpler equation $d(x-y)/dt=2(x-y)$, which has the solution $x-y=(x_0-y_0)e^{2t}$. If $x_0\neq y_0$, the right hand side is never zero and so $x=y=0$ can never occur.

For the second part, $x_0=y_0$. In this case, $x(t)=y(t)$ and so every solution is a parametrization of the line $x=y$. We can attempt to get a solution of the form $x=x_0-at$, $y=y_0-at$. This will be a solution if $u(t)=2(x_0-at)-a$. By taking $a=x_0/t_0$, $x=y=0$ at $t=t_0$.

A-6 (24). Let $\bar{y}=\lim_{n\to\infty}y_n$ and set $\bar{x}=\bar{y}/3$. We will show that $\bar{x}=\lim_{n\to\infty}x_n$. For any $\epsilon>0$ there is an N such that for all $n>N$, $|y_n-\bar{y}|<\epsilon/2$.

$$\begin{aligned}\epsilon/2 > |y_n-\bar{y}| &= |x_{n-1}+2x_n-3\bar{x}| = |2(x_n-\bar{x})+(x_{n-1}-\bar{x})| \\ &\geqq 2|x_n-\bar{x}| - |x_{n-1}-\bar{x}|.\end{aligned}$$

This may be rewritten as $|x_n-\bar{x}|<\epsilon/4+\frac{1}{2}|x_{n-1}-\bar{x}|$, which can be iterated to give

$$|x_{n+m}-\bar{x}| < \epsilon/4\left(\sum_{i=0}^{m}2^{-i}\right)+2^{-(m+1)}|x_{n-1}-\bar{x}| < \epsilon/2+2^{-(m+1)}|x_{n-1}-\bar{x}|.$$

By taking m large enough, $2^{-(m+1)}|x_{n-1}-\bar{x}|<\epsilon/2$. Thus for all sufficiently large k, $|x_k-\bar{x}|<\epsilon$.

SOLUTIONS. PART B

B-1 (229). The condition $24|n+1$ is equivalent to $n\equiv -1 \pmod 3$ and $n\equiv -1 \pmod 8$. Let d be a divisor of n, then $d\equiv 1$ or $2 \pmod 3$ and $d\equiv 1, 3, 5$ or $7 \pmod 8$. Since $d(n/d)=n\equiv -1 \pmod 3$ or $\pmod 8$, the only possibilities are:

$$\begin{array}{llll} d\equiv 1, & n/d\equiv 2 \pmod 3 & \text{or} & \text{vice versa} \\ d\equiv 1, & n/d\equiv 7 \pmod 8 & \text{or} & \text{vice versa} \\ d\equiv 3, & n/d\equiv 5 \pmod 8 & \text{or} & \text{vice versa.}\end{array}$$

In every case, $d+n/d\equiv 0 \pmod 3$ and $\pmod 8$. Thus $d+n/d$ is a multiple of 24. Note that $d\not\equiv n/d$ and thus no divisor is used twice in the pairing, so the sum of all the divisors is a multiple of 24.

B-2 (376). The number of elements in a subgroup is a divisor of the order of the group. Thus a proper subgroup can have no more than half of all the elements. Two subgroups always have the identity in common and hence their union cannot be the entire group.

An example for the second part of the problem is the Klein group, which has an identity and three elements x, y, z of order two. The product of any two distinct elements from $\{x, y, z\}$ is the third. This group is the union of three proper subgroups.

Alternate Solution: Let $G=H\cup K$, with H and K proper subgroups. There exists $k\in K$ with $k\notin H$. None of the elements in kH are in H and so $kH\subset K$. Hence $H\subset k^{-1}K=K$ and $K=H\cup K=G$, a contradiction.

B-3 (78). The first relation implies that

$$T_n = \frac{(n-1)(n-3)\cdots 3}{(n-2)(n-4)\cdots 2}\cdot\frac{1}{T_1} \qquad \text{if } n \text{ is even,}$$

$$T_n = \frac{(n-1)(n-3)\cdots 2}{(n-2)(n-4)\cdots 1}\cdot T_1 \qquad \text{if } n \text{ is odd.}$$

If n is odd,

$$\frac{T_n}{T_{n+1}} = (T_1)^2\cdot\frac{2}{1}\ \frac{2}{3}\ \frac{4}{3}\ \frac{4}{5}\ \frac{6}{5}\cdots\frac{(n-1)}{(n-2)}\ \frac{(n-1)}{n}$$

The Wallis product is $\pi/2=\frac{2}{1}\frac{2}{3}\frac{4}{3}\frac{4}{5}\frac{6}{5}\cdots$. After an even number of factors the partial product is less than $\pi/2$ and after an odd number of factors the partial product is greater than $\pi/2$. Thus for the case when n is odd, $T_n/T_{n+1}<\frac{1}{2}\pi T_1^2$. A similar calculation shows that, when n is even, $T_n/T_{n+1}<2/\pi T_1^2$. Since the limit of $T_n/T_{n+1}=1$, 1 is less than or equal to both $\frac{1}{2}\pi T_1^2$ and its reciprocal. This implies that $\pi T_1^2=2$.

B-4 (5). Place the curve so that its endpoints lie on the x-axis. Then take the smallest rectangle with sides parallel to the axes which covers the curve. Let its horizontal and vertical dimensions be a and b respectively. Let P_0 and P_5 be the endpoints of the curve, and let P_1, P_2, P_3, and P_4 be the points on the curve, in the order named, which lie one on each of the four sides of the rectangle. Draw the broken line $P_0P_1P_2P_3P_4P_5$. This line has length at most one. The horizontal components of the segments of this broken line add up at least to a, since one of the vertices of the broken line lies on the left end of the rectangle and one on the right end. The vertical segments add to at least $2b$ since we start and finish on the x-axis and go to both the top and bottom sides. This implies that the total length of the broken line is at least $(a^2+4b^2)^{1/2}$.

We now have that a and b both lie between 0 and 1 and that $a^2+4b^2\leqq 1$. Under these conditions the product ab is a maximum for $a=\frac{1}{2}\sqrt{2}$, $b=\frac{1}{4}\sqrt{2}$ and so the maximum of ab is $\frac{1}{4}$. Thus the area of the rectangle we have constructed is at most $\frac{1}{4}$.

B-5 (50). The following proof shows it is not necessary to stipulate that the a_n be integers. Suppose for some $\epsilon>0$ there are $x_j\to\infty$ with $k(x_j)/x_j\geqq\epsilon$. Note that if $1\leqq n\leqq k(x_j)$, then (because the a_n increase) $a_n\leqq a_{k(x_j)}\leqq x_j$ and $1/a_n\geqq 1/x_j$. Now for any positive integer N,

$$\sum_{n=N}^{\infty} 1/a_n \geqq \sup_j \sum_{n=N}^{k(x_j)} 1/a_n \geqq \sup_j \frac{k(x_j)-N}{x_j} \geqq \sup_j(\epsilon - N/x_j) = \epsilon.$$

But this contradicts the convergence of $\sum_{n=1}^{\infty} 1/a_n$, which implies

$$\lim_{N\to\infty}\sum_{n=N}^{\infty} 1/a_n = 0.$$

Comment: The above solution was formulated by a grader on the basis of a student's solution.

B-6 (6). Observe that $ABAB=9AB$. AB is of rank two so A is onto and B is one-to-one. Hence there exist matrices A' and B' such that $A'A=I=BB'$, where I is the 2×2 identity matrix. Then $A'(ABAB)B'=BA=9I$.

Alternate Solution: $(AB)^2=9AB$. The rank of BA is greater than or equal to the rank of $A(BA)B$, which is 2. Thus BA is nonsingular. But $(BA)^3=B(AB)^2A=B(9AB)A=9(BA)^2$ and the result follows since BA has an inverse.

Acknowledgments

The Director would like to acknowledge the assistance of the Questions Committee and the graders in preparing the above solutions and acknowledge the services of the following persons, who were graders for the competition: D. M. Bloom, L. R. Bragg, R. C. Busby, J. W. Dettman, R. A. DeVore, R. M. Dudley, D. J. Eustice, J. Froemke, R. A. Gambill, N. R. Grabois, M. Hausner, L. M. Kelly, M. S. Klamkin, W. S. Loud, D. G. Malm, J. M. Martin, R. Pollack, D. Rosen, P. J. Sally, M. E. Shanks, I. Schochetman, E. T. Wong.

CORRECTION TO "LIFE WITHOUT T_2"

ALBERT WILANSKY, Lehigh University

It was pointed out to me by D. E. Sanderson and J. W. Taylor that part of Example 2.2 (this MONTHLY, 77(1970)157–161) is not correct. Sanderson supplies the following example of a Baire space with a retraction onto a non-Baire subspace. Let I be the integers. Subsets of the plane are defined thus: let $S=\{(m/2^n, 0) : m, n\in I\}$, $Z_n=\{(x, 1/n) : x\neq m/2^n \text{ for all } m\in I\}$, $Z=\bigcup\{Z_n:n\in I\}$. The $Z\cup S$ is a Baire space, S is not, and $r:Z\cup S\to S$ is a retraction onto S, where r maps each component of Z onto the midpoint of its projection on the X-axis.

The following substitute example was suggested by G. A. Stengle. Let P be the property of certain topological spaces X: "X is homeomorphic with an interval $[a, b]$, $a\leqq b$". Then P is preserved under retraction but not under taking continuous images or closed subspaces.

MATHEMATICAL NOTES

EDITED BY DAVID DRASIN

Manuscripts for this Department should be sent to David Drasin, Division of Mathematical Sciences, Purdue University, Lafayette, IN 47907.

A BOREL SET WHICH CONTAINS NO RECTANGLES

RICHARD DARST AND CASPER GOFFMAN, Purdue University

This note contains an elementary example of a set E of positive two dimensional Lebesgue measure m_2 in the plane which not only contains no measurable rectangle of positive measure, but also has the property that every measurable rectangle of positive measure intersects the complement B of E in a set of posi-

tive measure. This latter property implies that if E is augmented by a set X of planar measure zero to form a measurable set Y, then Y will contain no measurable rectangle of positive measure.

Recall that a measurable rectangle of positive measure is a set of the form $U\times V$, where each of U and V is a measurable subset of the set R of real numbers, and the one dimensional Lebesgue measure m_1 of both U and V is positive.

Let A be a measurable subset of R such that both A and its complement $R-A$ have subsets of positive measure in every interval [1, p. 195, Ex. 4.1]. Denote by B the measurable subset of R^2 whose points belong to the lines with slope one which intersect the x-axis in a point of A, and let $E=R^2-B$.

Suppose that $C\times D$ is a measurable rectangle of positive measure. Then there are measurable subsets F and G of C and D and a set N of planar measure zero such that

(1) every point of F is a point of density of F,

(2) every point of G is a point of density of G, and

(3) $C\times D=(E\times F)\cup N$.

(See [1, p. 174, Theorem 2] and [2, p. 129].) Without loss of generality, we assume that $(0, 0)\in F\times G$.

Let U be an interval of length ϵ, symmetric about zero, and such that

$$m_1(U\cap F) > .9\epsilon \quad\text{and}\quad m_1(U\cap G) > .9\epsilon.$$

Let ϕ be defined on R by $\phi(t)=m_1([U\cap F]\cap[x;\ x=y+t$ for some $y\in(U\cap G)])$.

Then $\phi(0)=m_1([U\cap F]\cap[U\cap G])>.8\epsilon$. The function ϕ is continuous since if $T_t=[x+t:x\in T]$ then for S and T measurable sets, $m_1(S\cap T_t)$ is continuous in t. This is obvious since it clearly holds if S and T are intervals, and measurable sets can be approximated by finite unions of intervals.

Let δ be a positive number such that $\phi(t)\geqq .8\epsilon$, whenever $|t|\leqq\delta$. Then for $|t|\leqq\delta$, $m_1([x\in(U\cap F);\ x-y=t]$ for some $y\in(U\cap G))\geqq .8\epsilon$. Thus

$$m_2(B\cap[F\times G])\geqq\sqrt{2}\,(.8\epsilon)m_1(A\cap[-\delta,\delta])>0.$$

References

1. Casper Goffman, Real Functions, Prindle, Weber and Schmidt, Boston, 1967.
2. S. Saks, Theory of the Integral, translated by L. C. Young, Warszawa-Lwow, 1937.

A CONTINUOUS ALMOST PERIODIC FUNCTION HAS EVERY CHORD

M. C. TEWS, College of the Holy Cross

A continuous real-valued function f with domain the set of real numbers is said to be *almost periodic* if, for each positive number ϵ, the set

$$E\{\epsilon, f\} = \{t: |f(x+t)-f(x)| < \epsilon \text{ for all } x\}$$

is relatively dense, i.e., there is a positive number m such that $E\{\epsilon, f\}$ has at least one element in each open interval of length m. That a continuous periodic

function with period p is almost periodic follows by choosing m equal to $p+1$, since $E\{\epsilon, f\} \supset \{np : n \text{ an integer}\}$ for all $\epsilon > 0$. It is the purpose of this note to show that the technique used in [2] to prove that a continuous periodic function has horizontal chords of arbitrary length will yield the same result for almost periodic functions.

Before proceeding to the proof, we recall some facts about almost periodic functions, all of which are stated and proved in Chapter 1, section 1 of [1].

(i) *$t \in E\{\epsilon, f\}$ if and only if $-t \in E\{\epsilon, f\}$.*

(ii) *If $0 < q \leq m$, then $E\{q, f\} \subset E\{m, f\}$.*

(iii) *Almost periodic functions are uniformly continuous.*

(iv) *If f and g are almost periodic, then $E\{\epsilon, f\} \cap E\{\epsilon, g\}$ is relatively dense for each $\epsilon > 0$.*

The notation (m, n) is used for the interval $\{x : m < x < n\}$.

Let f be an almost periodic function and let a be a positive real number. To prove that f has a chord of length a, we shall show that $f(x+a) - f(x) = 0$ for some real number x. If this is not the case, then the continuous function defined by $x \to f(x+a) - f(x)$ is either always positive or always negative, by the intermediate value theorem. We assume, without loss of generality, that $f(x+a) - f(x) > 0$ for all x, and shall obtain a contradiction. Let $\epsilon > 0$. By (iii), we can choose a positive number δ such that

$$|f(x) - f(y)| < \epsilon/2, \qquad \text{whenever } |x - y| < \delta.$$

The function g, with period a, defined by $g(x) = \sin(2\pi x/a)$ is used as a tool. Let p be any positive number less than the minimum of δ and $a/8$ and let $\beta = g(p)$. Suppose t is such that

$$|g(x+t) - g(x)| < \beta \qquad \text{for all } x.$$

Choosing $x = 0$, this inequality gives

$$\left|\sin\left(\frac{2\pi t}{a}\right)\right| < \beta = \sin\left(\frac{2\pi}{a} p\right),$$

hence

$$t \in \bigcup_{-\infty < n < +\infty} \left(\frac{na}{2} - p, \frac{na}{2} + p\right);$$

however, choosing $x = a/4$ gives

$$\left|\sin\left[\frac{2\pi}{a}\left(\frac{a}{4} + t\right)\right] - \sin\left(\frac{2\pi}{a} \cdot \frac{a}{4}\right)\right| < \sin\frac{2\pi}{a} p,$$

which is the same as $|\cos(2\pi t/a) - 1| < \sin(2\pi p/a)$. Hence

$$\cos\left(\frac{2\pi}{a} t\right) > 1 - \sin\frac{2\pi p}{a} > 0,$$

so that

$$t \in \bigcup_{-\infty<n<+\infty} \left(na - \frac{a}{4},\ na + \frac{a}{4}\right).$$

Since p is less than $a/8$, intersecting these two sets yields

$$t \in \bigcup_{-\infty<n<+\infty} (na - p,\ na + p)$$

and, since p is less than δ, $t \in \bigcup_{-\infty<n<+\infty} (na-\delta,\ na+\delta)$. Therefore,

$$E\{\beta, g\} \subset \bigcup_{-\infty<n<+\infty} (na - \delta,\ na + \delta).$$

Finally, let $\eta = \min\ \{\beta, \epsilon/2\}$. We have proved that

$$E\{\eta, g\} \cap E\{\eta, f\} \subset E\{\eta, g\} \subset E\{\beta, g\} \subset \bigcup_{-\infty<n<+\infty} (na - \delta,\ na + \delta),$$

where the second set containment is by (ii). By (iv), the set on the left is relatively dense, and hence contains arbitrarily large numbers. We can, therefore, choose $t \in E\{\eta, g\} \cap E\{\eta, f\}$ such that $t = ma + q$, where m is a positive integer and $|t - ma| = |q| < \delta$. By the choice of δ, for all x we have

$$f(x + ma - t) - f(x) < \epsilon/2$$

and, since by (i), $-t$ is also a member of $E\{\eta, f\}$, we have

$$f(x + ma) - f(x + ma - t) < \eta.$$

Recalling that $x \to f(x+a) - f(x)$ is positive, we obtain

$$\begin{aligned} 0 < f(x + a) - f(x) &\leqq \sum_{k=1}^{m} [f(x + ka) - f(x + (k - 1)a)] \\ &= f(x + ma) - f(x) = [f(x + ma) - f(x + ma - t)] \\ &\quad + [f(x + ma - t) - f(x)] < \eta + \epsilon/2 \leqq \epsilon. \end{aligned}$$

However, since ϵ is arbitrary, this implies that $f(x+a) - f(x)$ is zero for all x, which is our desired contradiction. We conclude that f has a horizontal chord of length a.

In [2], the "Dirichlet box principle" is used as a basic tool. That tool is used in this note but is hidden in the proof of 10° of [1] which is used to prove (iv).

References

1. A. S. Besicovitch, Almost Periodic Functions, Dover, New York, 1954.

2. J. B. Diaz and F. T. Metcalf, A continuous periodic function has every chord twice, this Monthly, 74 (1967) 833–835.

A CHARACTERIZATION AND A CANONICAL DECOMPOSITION OF HURWITZIAN MATRICES

R. DATKO, Georgetown University, and V. SESHADRI, McGill University

A complex $n\times n$ matrix A is called *Hurwitzian* if all its eigenvalues have real parts negative. In this note we shall use a result of A. M. Lyapunov to obtain a characterization and a canonical decomposition for such matrices. First some notation will be established.

NOTATION: (a) $A, B, \cdots$, etc. will denote complex $n\times n$ matrices. (b) A^* will denote the conjugate transpose of a matrix A. (c) If Q is a positive definite Hermitian matrix, then we shall denote this by $Q>0$. (d) If $Q>0$ then $Q^{1/2}$ will denote the unique positive definite square root of Q (see, e.g., [1] p. 92).

The basis for the results obtained in this note is the following lemma (see, e.g., [2] p. 155):

LEMMA. *If A is Hurwitzian, the solution in B of the equation $A^*B+BA=-C$ is unique and is given by the formula*

$$B = \int_0^\infty e^{A^*t}Ce^{At}dt.$$

If $C>0$, then $B>0$. Moreover, if $C>0$ and $B>0$ and $A^*B+BA=-C$, then A is Hurwitzian.

Let A be Hurwitzian and choose $C=I$. Then on the basis of the lemma, there is a unique $B>0$ such that

$$A^*B + BA = -I. \tag{1}$$

Define the matrix S by the relation

$$-\tfrac{1}{2}B^{-1} + S = A. \tag{2}$$

Substituting this expression for A in (1) yields the equation $S^*B+BS=0$ or, what is the same,

$$-S^*B = BS. \tag{3}$$

We multiply both sides of (3) by $B^{-1/2}$ on the left and then by $B^{-1/2}$ on the right to obtain

$$-B^{-1/2}S^*B^{1/2} = B^{1/2}SB^{-1/2} \tag{4}$$

which shows that S is similar to a skew-Hermitian matrix.

PROPERTY 1. *Any Hurwitzian matrix A can be written in the form*

$$A = -Q + T,$$

where $Q>0$ and T is similar to a skew-Hermitian matrix.

Proof. Let A be a Hurwitzian matrix and let B satisfy (1). Using equation (2) we write $A=-\frac{1}{2}B^{-1}+S$. By equation (4), S is similar to a skew-Hermitian matrix and since $B>0$, $B^{-1/2}/2>0$. Letting $Q=B^{-1}/2$ and $T=S$ establishes Property 1.

PROPERTY 2. *A matrix A is Hurwitzian if and only if it is similar to a matrix of the form*

$$-Q+T,$$

where $Q>0$ and T is skew-Hermitian.

Proof. If A is Hurwitzian and B satisfies (1) then, by equation (2), $A=\frac{1}{2}B+S$. Multiplying both sides of this last equation by $B^{1/2}$ on the left and then by $B^{-1/2}$ on the right we obtain

$$B^{1/2}AB^{-1/2}=(B^{-1}/2)+B^{1/2}SB^{-1/2}. \tag{5}$$

Letting $Q=B^{-1}/2$ and $T=B^{1/2}SB^{-1/2}$ and taking cognizance of equation (4) establishes the necessity of the property.

Next, suppose A is similar to $K=-Q+T$ where $Q>0$ and T is skew-Hermitian. Then $K+K^*=-2Q$. Hence K satisfies the hypotheses of the lemma where $B=I>0$ and $C=2Q>0$ so that K is Hurwitzian. From this it follows that A is Hurwitzian.

Property 1 is the canonical decomposition mentioned above and Property 2 is the characterization.

References

1. R. Bellman, Introduction to Matrix Theory, McGraw-Hill, New York, 1960.
2. A. Halanay, Differential Equations, Academic Press, New York and London, 1966.

ON GROUPS OF MOTIONS GENERATED BY TWO ROTATIONS

LEOPOLD FLATTO, Yeshiva University

1. Let R_1 and R_2 denote respectively the rotations of the Euclidean plane about the distinct points P_1 and P_2, the angle of rotation being $2\pi/n$, where n is a positive integer. It is a well-known fact that the group G generated by R_1 and R_2 is discrete if and only if $n=1, 2, 3, 4, 6$. A simple geometric proof of this fact was given by Barlow (see [1], page 60). In this note we give an algebraic proof which actually yields somewhat more information. In section 2 we give an application of Theorem 1 to the study of level curves of harmonic functions in two variables.

THEOREM 1. *Let G be the group generated by two distinct two-dimensional rotations, the angle of rotation being $2\pi/n$. If $n\neq 1, 2, 3, 4, 6$ then the subgroup G_0 consisting of the translations in G is dense in the group G_T of all translations.*

Proof. We may assume that the centers of rotations of R_1 and R_2 in the complex plane are given respectively by 0 and 1. Thus

$$R_1(z) = \zeta z, \qquad R_2(z) = \zeta z + (1 - \zeta),$$

where $\zeta = \exp[2\pi i/n]$. Let $T_m(z) = R_1^m R_2 R_1^{-(m+1)}(z)$. Direct computations yield $T_m(z) = z + \zeta^m(1-\zeta)$.

G_0 contains the translations

$$T_0^j T_1^k T_{n-1}^k(z) = z + (j + kx)(1 - \zeta),$$

$$T_1^j T_2^k T_0^k(z) = z + (j + kx)\zeta(1 - \zeta),$$

where $x = \zeta + \zeta^{-1} = 2\mathrm{Re}\,\zeta$ and j, k are arbitrary integers.

For $n > 2$, $T_0^j T_1^k T_{n-1}^k$ and $T_1^j T_2^k T_0^{-k}$ are translations in two different directions. If x is irrational, then Kronecker's theorem [2, page 370] states that the numbers $\{j + kx\}$ $(-\infty < j, k < \infty)$ form a dense set of real numbers. Thus G_0 is dense in G_T provided x is irrational. If x is rational then $\zeta^2 - x\zeta + 1 = 0$ implies that the minimal polynomial of ζ over the rational field has degree $\leqq 2$. But the minimal polynomial of ζ over the rational field is the cyclotomic polynomial of degree $\phi(n)$ [3, page 162]. Thus $\phi(n) \leqq 2$, so that $n = 1, 2, 3, 4, 6$.

We have thus shown that G_0 is dense in G_T except for $n = 1, 2, 3, 4, 6$. These latter values are indeed exceptional as G is then known to be discrete [1].

2. We now give an application of the previous discussion. We employ the following terminology. A curve Γ is said to be a *level curve of a harmonic function* if there exists a function harmonic in the entire plane, except for isolated singularities, which vanishes on Γ but not identically. We obtain the following result:

THEOREM 2. *Let Γ be the image of the line $x = c$ $(c \neq 0)$ under the map $w = z^n$, $z = x + iy$. Then Γ is a level curve of a harmonic function iff $n = 1, 2, 3, 4, 6$.*

Proof: Let $v(z) = u(z^n)$. If $u(z)$ is a harmonic function vanishing on Γ, then $v(z)$ is harmonic and vanishes on the line $x = c$. It follows from the reflection principle for harmonic functions that $v(z) = -v(R(z))$ where R denotes the reflection in the line $x = c$. Let $R_1(z) = \zeta z$ and $R_2(z) = RR_1^{-1}R(z)$, $\zeta = \exp[2\pi i/n]$. Then $v(R_1(z)) = v(R_2(z)) = v(z)$ so that $v(g(z)) = v(z)$ for $g \in G$, G being the group generated by R_1 and R_2. R_1 and R_2 denote respectively rotations by an angle $2\pi/n$ about the points 0, $2c$. It follows from Theorem 1 that for $n \neq 1, 2, 3, 4, 6$ $v(z)$ is a constant. Since $v(z) = 0$ on Γ, $u(z) = v(z) = 0$ for all z. Hence Γ is not a level curve for these values of n.

For $n = 1, 2, 3, 4, 6$, Γ is a level curve of a harmonic function. This may be shown as follows. We establish the existence of a nonconstant function $f(z)$ with the following properties:

1. $f(z)$ is meromorphic in the entire plane.
2. $f(\zeta z) = f(z)$ for all z, where $\zeta = e^{2\pi i/n}$.
3. $\mathrm{Im}[f(z)] = 0$ on the line $x = c$.

It follows that $f(z^{1/n})$ is well defined and meromorphic in the entire plane. Furthermore, $\mathrm{Im}[f(z^{1/n})] = 0$ for $z \in \Gamma$. The nonconstant harmonic function $u(z)$

$=\operatorname{Im}[f(z^{1/n})]$ has only isolated singularities, as $f(z^{1/n})$ is meromorphic in the entire plane, and Γ is a level curve of $u(z)$.

It remains to establish the existence of $f(z)$ having the properties 1–3. Since the existence of an $f(z)$ for a given n implies the existence of an $f(z)$ for any divisor of n, it suffices to consider the cases $n=4$, 6. Suppose that $n=4$. Let Ω denote the lattice consisting of the points $mc+nci$, where $-\infty<m<\infty$, $-\infty<n<\infty$, and denote the points in Ω by ω. Let $f(z)=\sum_\omega 1/(z-\omega)^4$, where the summation extends over all $\omega\in\Omega$. $f(z)$ is a nonconstant elliptic function with the periods c and ci. We observe that as ω varies over Ω, so do $-\bar{\omega}$ and $-i\omega$. It follows that

$$f(iz) = \sum_\omega \frac{1}{(iz-\omega)^4} = \sum_\omega \frac{1}{(z+i\omega)^4} = \sum_\omega \frac{1}{(z-\omega)^4} = f(z) \tag{3.1}$$

$$\overline{f(iy)} = \sum_\omega \frac{1}{(-iy-\bar{\omega})^4} = \sum_\omega \frac{1}{(iy+\bar{\omega})^4} = \sum_\omega \frac{1}{(iy-\omega)^4} = f(iy). \tag{3.2}$$

Thus $f(z)$ is real for $z=iy$. Since $f(z)$ has period c, $f(z)$ is real for $z=c+iy$. Thus $f(z)$ satisfies properties 1–3.

The case $n=6$ is treated in a similar fashion. Let Ω denote the lattice consisting of the points $mc+nc\zeta$ where $-\infty<m<\infty$, $-\infty<n<\infty$. Let

$$f(z) = \sum_\omega \frac{1}{(z-\omega)^6};$$

$f(z)$ is a nonconstant elliptic function with the periods c, $c\zeta$. We observe that as ω varies over Ω, so do $-\bar{\omega}$ and $\zeta^{-1}\omega$. It follows that

$$f(\zeta z) = \sum_\omega \frac{1}{(\zeta z-\omega)^6} = \sum_\omega \frac{1}{(z-\zeta^{-1}\omega)^6} = \sum_\omega \frac{1}{(z-\omega)^6} = f(z), \tag{3.3}$$

$$\overline{f(iy)} = \sum_\omega \frac{1}{(-iy-\bar{\omega})^6} = \sum_\omega \frac{1}{(iy+\bar{\omega})^6} = \sum_\omega \frac{1}{(iy-\omega)^6} = f(iy). \tag{3.4}$$

Thus $f(z)$ is real for $z=iy$. Since $f(z)$ has period c, $f(c+iy)$ is real. Thus $f(z)$ satisfies properties 1–3.

Finally, we remark that if $n=1$ or 2, then Γ is a level curve of a nonconstant harmonic function without singularities. For we may choose $f(z)$ in these cases to be respectively $i(z-c)$, $\cos \pi z/c$. This is no longer true for $n=3$, 4, 6. In the latter cases the curve Γ contains a closed loop. Thus if $u(z)$ is an everywhere harmonic function vanishing on Γ, then $u(z)$ will vanish inside the loop and hence identically.

References

1. H. S. M. Coxeter, Introduction to Geometry, Wiley, New York, 1962.
2. G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, Oxford Clarendon Press, 1945.
3. B. L. van der Waerden, Modern Algebra, vol. I, Ungar, New York, 1950.

ON THE LOCATION OF SINGULARITIES OF ANALYTIC FUNCTIONS

A. R. BLASS, Harvard University, and C. V. STANOJEVIC, University of Missouri at Rolla

Consider an analytic function given by its Taylor expansion about a point which, for convenience, we take to be the origin

$$f(z) = \sum_{n=0}^{\infty} a_n z^n. \tag{1}$$

If this series has a finite radius of convergence, then there is at least one singular point of f on the circle of convergence [4, Sec. 10.2]. In principle, the location of singular points is determined by the coefficients a_n; thus, the problem arises of actually finding the singularities in terms of the a_n. A theorem due to Pringsheim [5], [4, Sec. 5.7] asserts that, if all but finitely many of the a_n are real and non-negative, or even satisfy only $|\arg a_n| \leqq \beta$ for some $\beta < \pi/2$, [3], then the positive real point on the circle of convergence is singular. A rather deep theorem of Fabry [1], [2] asserts that, if a_n/a_{n+1} approaches a limit, a, (so, by the ratio test, $|a|$ is the radius of convergence) then a is a singularity.

In this paper, we shall prove, in an elementary way, a theorem related to that of Fabry. The hypothesis of our theorem assumes something more than convergence of a_n/a_{n+1}, but our conclusion also asserts more than the conclusion of Fabry's theorem.

Before stating our theorem, we must introduce some definitions and notations. Suppose z is a fixed point inside the circle of convergence of (1). Then (1) may be differentiated term by term as often as we want. Let us differentiate it k times and then multiply the result by $z^k/k!$. The resulting series

$$\sum_{n=k}^{\infty} a_n \binom{n}{k} z^n \tag{2}$$

converges absolutely. In particular, its partial sums are bounded. If z is not a zero of $f^{(k)}(z)$, then we can find a number M such that

$$\left| \sum_{n=k}^{N} a_n \binom{n}{k} z^n \right| \leqq M \left| \sum_{n=k}^{\infty} a_n \binom{n}{k} z^n \right| \qquad \text{for all } N \geqq k. \tag{3}$$

In general, M will depend on z and on k. If for some value of $z \neq 0$, it happens that M may be chosen independent of k, then we call such a value of z a *special* point for the function f.

For brevity, we shall write "the circle about p" instead of "the circle of convergence of the Taylor series of $f(z)$ about p."

Let us remark that the hypothesis of Fabry's theorem says precisely that the series

$$\sum_{n=1}^{\infty} \left(\frac{a_n}{a_{n+1}} - \frac{a_{n-1}}{a_n} \right) = a - \frac{a_0}{a_1} \tag{4}$$

converges. Now we are in a position to state our theorem.

THEOREM. *Let an analytic function f have the expansion* (1), *and suppose none of the a_n are zero and the series* (4) *converges to $a-(a_0/a_1)$ and converges absolutely. If $z\neq 0$ is a special point for f, then the circle about z passes through a.*

Proof. Let us introduce the notations

$$S_k(N)=\sum_{n=k}^{N}\binom{n}{k}a_nz^n,\qquad S_k=\sum_{n=k}^{\infty}\binom{n}{k}a_nz^n=\lim_{N\to\infty}S_k(N).$$

Applying the ratio test to the expansion

$$f(w)=\sum_{k=0}^{\infty}\left(\sum_{n=k}^{\infty}a_n\binom{n}{k}z^{n-k}\right)(w-z)^k=\sum_{k=0}^{\infty}\frac{S_n}{z^k}(w-z)^k$$

we see that the circle about z has radius $\lim_k |z(S_k/S_{k+1})|$ provided that this limit exists. We shall show that it does exist and equals $|a-z|$. Observe that, since z is special, we have

$$|S_k(N)|\leqq M|S_k|\qquad\text{for all } N \quad\text{and}\quad k\leqq N. \tag{3'}$$

We calculate:

$$\begin{aligned}
&z(S_k(N-1)+S_{k+1}(N-1))-aS_{k+1}(N)\\
&=z\left(\sum_{n=k}^{N-1}\binom{n}{k}a_nz^n+\sum_{n=k+1}^{N-1}\binom{n}{k+1}a_nz^n\right)-aS_{k+1}(N)\\
&=\sum_{n=k}^{N-1}\binom{n+1}{k+1}a_nz^{n+1}-a\sum_{n=k+1}^{N}\binom{n}{k+1}a_nz^n\\
&=\sum_{n=k+1}^{N}\binom{n}{k+1}\left(\frac{a_{n-1}}{a_n}-a\right)a_nz^n\\
&=\sum_{n=k+1}^{N-1}\left(\frac{a_{n-1}}{a_n}-\frac{a_n}{a_{n+1}}\right)\sum_{j=k+1}^{n}\binom{j}{k+1}a_jz^j+\left(\frac{a_{N-1}}{a_N}-a\right)\sum_{j=k+1}^{N}\binom{j}{k+1}z^ja_j\\
&=\sum_{n=k+1}^{N-1}\left(\frac{a_{n-1}}{a_n}-\frac{a_n}{a_{n+1}}\right)S_{k+1}(n)+\left(\frac{a_{N-1}}{a_N}-a\right)S_{k+1}(N).
\end{aligned}$$

If we let $N\to\infty$, the last term vanishes because $a=\lim_N a_{N-1}/a_N$, and $S_{k+1}(N)$ is bounded, so we get

$$\begin{aligned}
|zS_k-(a-z)S_{k+1}|&=\left|\sum_{n=k+1}^{\infty}\left(\frac{a_{n-1}}{a_n}-\frac{a_n}{a_{n+1}}\right)S_{k+1}(n)\right|\\
&\leqq M|S_{k+1}|\sum_{n=k+1}^{\infty}\left|\frac{a_{n-1}}{a_n}-\frac{a_n}{a_{n+1}}\right|,
\end{aligned}$$

by (3′). Thus

$$\left| z\frac{S_k}{S_{k+1}} - (a - z) \right| \leqq M \sum_{n=k+1}^{\infty} \left| \frac{a_{n-1}}{a_n} - \frac{a_n}{a_{n+1}} \right| = o(1), \qquad k \to \infty,$$

because (4) converges absolutely. This calculation shows that the circle about z has radius exactly $|a - z|$, and therefore passes through a.

If we combine the theorem just proved with the fact that the circle about any point has a singularity on its circumference but none in its interior, and the obvious fact that the circle about O has radius a, we immediately find the following results:

If (4) converges absolutely, and if there is a special point z on the straight line segment joining O to a, then a is a singularity of f (as in Fabry's theorem). If there is a special point z on the line segment joining O to $-a$, then, not only is a a singularity, but it is the *only* singularity on the circle about O. If there are two special points, one on the line segment from O to ia, the other on the line segment from O to $-ia$, then either a or $-a$ is singular (in fact we know, by Fabry's Theorem, that a is singular), and no point other than these two on the circle about O is singular.

References

1. L. Bieberbach, Analytische Fortsetzung, Ergeb. der Math., **3** (1955) 64.
2. P. Dienes, The Taylor Series, Oxford, (1931) 377.
3. A. Dinghas, Vorlesungen über Funktionentheorie, Springer-Verlag, (1961) 194.
4. E. Hille, Analytic Function Theory, Blaisdell, New York, 1959.
5. A. Pringsheim, Über Funktionen welche in gewissen Punkten endliche Differentialquotienten jeder endlichen Ordnung, aber keine Taylor'sche Reihenentwicklung besitzen, Math. Ann., **44** (1894) 41–56.

RESEARCH PROBLEMS

EDITED BY VICTOR KLEE

In this Department the Monthly presents easily stated research problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Victor Klee, Department of Mathematics, University of Washington, Seattle, WA 98105.

ON SETS OF DISTANCES OF n POINTS

PAUL ERDÖS, Israel Institute of Technology, Haifa

Let $f(n)$ be the largest integer so that n distinct points in the plane always determine at least $f(n)$ distinct distances. It is easy to see that $f(3) = 1, f(4) = f(5) = 2$, $f(6) = f(7) = 3$. I proved [1]

(1) $$\sqrt{n-1}-1<f(n)<\frac{c_1 n}{\sqrt{\log n}}\cdot$$

Moser proved [2]

(2) $$f(n)>\frac{n^{2/3}}{2\,(9^{1/3})}-1$$

which is the best-known lower bound for $f(n)$.

It seems likely that $c_2n/(\log n)^{1/2}$ is the right order of magnitude for $f(n)$. In fact perhaps the following result holds: Let $x_1, \cdots, x_n$ be n distinct points in the plane. Then for at least one point x_i there are at least $c_3n/(\log n)^{1/2}$ distinct numbers amongst the distances $d(x_i, x_j)$, where $1\leqq j\leqq n$.

Assume next that the points $x_1, \cdots, x_n$ are vertices of a convex n-gon. I conjectured [1] and Altman [3] proved that the n points determine at least $[n/2]$ distinct distances. (The regular n-gon shows that this is best possible.) I made two further conjectures [1]. Let $x_1, \cdots, x_n$ be the vertices of a convex n-gon. Then there always is an x_i so that there are at least $[n/2]$ distinct distances among the $d(x_i, x_j)$, where $1\leqq j\leqq n$ and $j\neq i$. This is probably true but has not yet been settled. The second conjecture asserts that there always is an x_i so that there are no three vertices equidistant from it.

The second conjecture would clearly imply the first, but Danzer disproved it (unpublished). In fact Danzer showed that for each k, there is a convex n-gon with $n>n_0(k)$ so that every vertex has at least k vertices which are equidistant from it.

Let $g(n)$ be the largest integer so that there are n points $x_1, \cdots, x_n$ in the plane for which there are $g(n)$ pairs x_i, x_j with $d(x_i, x_j)=1$. I proved [1]

$$n^{1+c/\log\log n}<g(n)<cn^{3/2}.$$

It seems likely that the lower bound gives the correct order of magnitude, but I could not even prove $g(n)=o(n^{3/2})$.

All these problems can be posed in the case the points are in k-dimensional Euclidean space E_k. Curiously some of them become more tractable for $k\geqq 4$ [4].

Let 7 points be given in E_2. L. M. Kelly and I proved [5] that there are always three of them which determine a nonisosceles triangle. The regular pentagon and its center shows that 7 is best possible. Croft [6] proved that 9 points in E_3 gives the best possible answer and believes that $2k+3$ points in E_k always determine the vertices of a nonisosceles triangle.

More generally one can ask the following question. Let $f(n, k)$ be the smallest integer so that if $x_1, \cdots, x_l$ are $l=f(n, k)$ points in E_k, one can always select k of them so that all the $C_{k,2}$ distances are distinct. A good estimation for $f(n, k)$ seems difficult even for $k=1$. I conjecture $f(n, 1)=(1+o(1))n^2$. A result of Turán and myself [7] shows that $f(n, 1)\geqq(1+o(1))n^2$.

For some of these and other geometric problems, see my Hungarian paper in Mat. Lapok, 8 (1957) 86–92.

References

1. P. Erdös, On sets of distances of n points, this MONTHLY, 53 (1946) 248–250.
2. L. Moser, On the different distances determined by n points, *ibid.*, 59 (1952) 85–91.
3. E. Altman, On a problem of P. Erdös, *ibid.*, 70 (1963) 148–157.
4. P. Erdös, On some applications of graph theory to geometry, Canad. Jour. Math., 19 (1967) 968–971. On sets of distances of n points in Euclidean space, Publ. Math. Inst. Hung. Acad. Sci., 5 (1960) 165–169.
5. ———, Elementary problem 735, this MONTHLY, 54 (1947) 227–229.
6. H. T. Croft, 9 point and 7 point configurations in 3-spaces, Proc. London Math. Soc., 12 (1962) 400–424.
7. P. Erdös and P. Turán, On a problem of Sidon, Jour. London Math. Soc., 16 (1941) 292–296.

CLASSROOM NOTES

EDITED BY DAVID DRASIN

Manuscripts for this Department should be sent to David Drasin, Division of Mathematical Sciences, Purdue University, Lafayette, IN 47907.

SIZING UP SETS AND CONTINUITY—DIFFERENTIABILITY RELATIONSHIPS

S. G. WAYMENT, Utah State University

1. Introduction. Let S be the closed interval $[0, 1]$. There are three standard methods of describing the "size" of a subset of S, i.e. (a) cardinality, (b) category, and (c) Lebesgue measure. The purpose of this paper is to demonstrate by examples the discrepancies which arise when comparing the sizes of sets with these methods. One of the resulting sets is used to construct a function which satisfies conditions "similar" to conditions under which no such function exists. More specifically, we construct a function which is continuous on a dense set D of measure zero and is discontinuous on $S-D$ which has measure one. We also ask an interesting question concerning the existence of a continuous function whose derivative exists only on a countable dense set. In the final section of the paper we give a function which is a counterexample to some results stated in [1] and [2] on the derivative of a function with a dense set of discontinuities.

2. Comparing some sets. Let M_n be a set consisting of n elements from S. Let H be the set of reciprocals of positive integers, let R be the set of rational points in S, let C be the Cantor set, and let P be the nonmeasurable set described in [3]. Finally, suppose $\langle r_i \rangle$ is an ordering of R and let I_{ik} be an open interval centered at r_i such that the length $l(I_{ik}) < 1/k2^i$. For $k \geqq 3$ let $O_k = \bigcup_{i=1}^{\infty} I_{ik}$. Then the measure of O_k satisfies $m(O_k) < 1/k$. Further, $\tilde{O}_k = S - O_k$ is nowhere dense and $\bigcup_{k=3}^{\infty} \tilde{O}_k$ is first category while $G_\delta = \bigcap_{k=3}^{\infty} O_k$ is of second category and $m(G_\delta) = 0$. If we assume the continuum hypothesis and recall that countability of $A \subset S$ implies that $m(A) = 0$ and A is of first category, then we can obtain an ordering relative to "size" by each of the three mentioned methods. We shall use $<$ and

= to represent this ordering under each method and shall insert parenthesis to indicate a rather loose classification as to "size."

(a) With respect to cardinality we have (finite) < (countably infinite) < (uncountably infinite) and

$$(\phi < M_1 < \cdots < M_n < \cdots) < (H = R) < (C = O_k = \tilde{O}_k = G_\delta = \tilde{G}_\delta = P = S).$$

(b) With respect to category we have (nowhere dense) < (first category) < (second category) and

$$(\phi = M_1 = \cdots = M_n = \cdots = H = C = \tilde{O}_k) < (R = \tilde{G}_\delta) < (O_k = G_\delta = P = S).$$

(c) With respect to measure we have (measure zero) < (measure between zero and one) < (measure one) and

$$(\phi = M_1 = \cdots = M_n = \cdots = H = R = C = G_\delta) < (O_k < \tilde{O}_k) < (\tilde{G}_\delta = S).$$

Note that P is not comparable with respect to measure.

The following relations are known to hold:

(i) S(which is uncountable) is not the union of a countable class of countable sets;

(ii) S (which is of second category) is not the union of a countable class of sets of first category; and

(iii) S (which has measure one) is not the union of a countable class of sets of measure zero.

It is of interest to note that $S = G_\delta \cup \tilde{G}_\delta$ is the union of two "small" sets, one of measure zero (second category) and one of first category (measure one). In view of mentioned results, no other such decomposition of S into "small" sets holds.

3. On the existence of certain functions. A standard example of a function which is continuous on the irrationals and is discontinuous on the rationals is $f(x) = 0$ when $x \in S - R$ and $f(x) = 1/q$ when $x = p/q \in R$. (In this and the next example p/q is understood to be lowest terms.) It is well known and easily proved, using category arguments, that there does not exist a function which is continuous on the irrationals and discontinuous on the rationals. We next establish the existence of a function which is "almost" of this nature.

THEOREM. *There exists a function f defined on $S = [0, 1]$ such that f is continuous on a dense set D of measure zero and is discontinuous on $S - D$.*

Proof. Let $D = G_\delta$ and let $f_k(x) = 0$ if $x \in O_k$ and $f_k(x) = 1/2^k$ if $x \in S - O_k$. Then $f = \sum_{k=3}^{\infty} f_k$ is the desired function. Note that if $x \in G_\delta$, then $f(x) = 0$ and if $x \notin G_\delta$ then $f(x) > 0$. Since $\overline{G}_\delta = S$, this shows f is discontinuous for $x \notin G_\delta$. If $x_0 \in G_\delta$ and $\epsilon > 0$, then there exists N such that $\sum_{i=N}^{\infty} 1/2^i < \epsilon$ and there exists an open interval I so that $x_0 \in I \subset O_i$ for $i = 1, 2, \cdots, N$. Hence we conclude that if $x \in I$ then

$$0 < f(x) - f(x_0) = f(x) < \sum_{i=N}^{\infty} 1/2^i < \epsilon$$

and f is continuous at x_0. Hence f is continuous on a dense set of measure zero and discontinuous elsewhere.

REMARK. If f is continuous on a dense set, then the set of points of continuity must be of second category. Hence there does not exist a function f continuous on $\tilde{G}_\delta$ and discontinuous on G_δ since $m(\tilde{G}_\delta)=1$ and hence $\tilde{G}_\delta$ is dense but of first category.

Another standard example of a function that is continuous on the irrationals and discontinuous on the rationals is given by $g(x)=\sum_{0\leq p/q\leq x} 1/q^3$ where p/q is rational. Since g is increasing, g is integrable and $G(x)=\int_0^x g$ is continuous for each x. Furthermore $G'(x)$ exists and is equal to $g(x)$ at points where g is continuous. Also, $G'(x)$ is easily seen not to exist at any rational and hence G is a continuous function whose derivative exists at each irrational and fails to exist at each rational. It is known that there exists a continuous function $h(x)$ on $[0, 1]$ such that $0<h(x)<1$ and the derivative fails to exist for each x. Hence $h(x)\prod_{i=1}^n (x-r_i)^2$ is a continuous function whose derivative exists only on the finite set $\{r_i\}$. This leads to the following question:

QUESTION. Does there exist a continuous function whose derivative exists at each rational and fails to exist at each irrational?

4. A Counterexample. On pages 126–7 of [1] a proof is given to show that if f is a function with a dense set of discontinuities, then f' exists only on a set of first category. However, the author of [1] divides an inequality by z_k-x_k and apparently assumes this quantity always to be nonnegative since the inequality is not reversed. The damage is irreparable as consideration of $f(x)=\sum_{0\leq r\leq x} 1/q^3$ illustrates. As already mentioned, f is continuous at each irrational in $[0, 1]$ and discontinuous at each rational in $[0, 1]$. Further, if x is an irrational number and N is an integer, then there exists a δ such that $(x-\delta, x+\delta)$ contains no rational with denominator (in lowest terms) less than N. We note also that the maximum number of rationals with denominator g in (x, y) is $g(y-x)$. Hence if $x<y<x+\delta$,

$$\Delta f = f(y) - f(x) = \sum_{x<r\leq y} 1/q^3 < \sum_{q\geq N} [q(y-x)]1/q^3 = (y-x)\sum_{q\geq N} 1/q^2.$$

Thus by choosing N sufficiently large (δ sufficiently small) we obtain

$$0 \leqq \Delta f/(y-x) < \epsilon$$

and since a similar argument is valid for $y<x$ we conclude $f'(x)=0$ for each irrational $x\in[0, 1]$. The set of irrationals in $[0, 1]$ is of second category. The proof in [1] actually shows that if the left hand limit or right hand limit fails to exist on a dense set, then f' can exist only a set of first category.

It is further stated in [1] and [2] that "if a function is discontinuous at the points of an everywhere dense set and differentiable (hence continuous) at the points of another everywhere dense set, then it must be continuous and not differentiable at the points of a set of second category." This result is also invalid since $f(x)=\sum_{0\leq r\leq x} 1/q^3$ is continuous and differentiable at each irrational.

References

1. R. P. Boas, Jr., A Primer of Real Functions, Carus Mathematical Monographs, Number 13, MAA; Wiley, New York, 1960.
2. M. K. Fort, Jr., A Theorem Concerning Functions Discontinuous on a Dense Set, this MONTHLY, 58 (1951) 408–410.
3. H. L. Royden, Real Analysis, 2nd ed., Macmillan, New York, 1968, pp. 63–64.

THE NUMBER OF IRREDUCIBLE POLYNOMIALS OF DEGREE n OVER GF(p)

G. J. SIMMONS, Sandia Laboratories and University of New Mexico

Many algebra texts [1, 2, for example] introduce the number theoretic Möbius inversion formula prior to discussing the cyclotomic extensions of a field, since it provides a simple means for computing the cyclotomic polynomials. It is apparently not well known, however, that the Möbius function can also be used to derive an expression for the number of irreducible monic polynomials of degree n over GF(p). Carmichael [3] gives a lengthy calculation of this quantity, which he calls N_n, by applying the inclusion-exclusion principle to the number of prime factors of the divisors of n. The following simple derivation of N_n can either be used in a course in number theory as an example at the time the Möbius inversion formula is introduced, or in a course in algebra as a supplement to the discussion of irreducible polynomials.

We include a brief derivation of the following theorem, which appears in most algebra texts [3, p. 258; 4, Problem 45.8, p. 369, etc.] and a statement of the Möbius inversion theorem to make this note self-contained.

THEOREM 1. *The polynomial $x^{p^n}-x$ is the product of all monic irreducible polynomials over* GF(p) *whose degree d divides n.*

Proof. Let $p(x)$ be a monic irreducible polynomial of degree d over GF(p). Then $p(x)$ can be used to extend GF(p) to a field K containing p^d elements. If $p(x)$ divides $x^{p^n}-x$ then K may be embedded in the splitting field GF(p^n) of $x^{p^n}-x$. Then GF(p^n) is a vector space of dimension e over K. Viewing both fields as vector spaces over GF(p) and counting dimensions, $de=n$. Conversely, if $de=n$, then K is isomorphic to the fixed field of GF(p^n) under the automorphism $x\to x^{p^e}$ since each $\alpha\in K$ satisfies $\alpha^{p^d}=\alpha$. Thus $p(x)$ has a root α in GF(p^n). But α is also a root of $x^{p^n}-x$, hence $p(x)$ divides $x^{p^n}-x$.

Since each monic irreducible polynomial of degree d has d elements of GF(p^n) as roots, and no element of GF(p^n) is a root of more than one irreducible polynomial, counting shows that each irreducible polynomial occurs at most once in the factorization of $x^{p^n}-x$. ∎

If N_d is the number of irreducible monic polynomials of degree d over GF(p), then the sum of the weighted degrees is given by

$$p^n = \sum_{d\mid n} dN_d \tag{1}$$

where d ranges over all divisors of n.

The *Möbius function*, $\mu(d)$, of an integer d with the prime factorization

$$d = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_k^{\alpha_k}$$

is defined by

$$\begin{aligned} \mu(1) &= 1 \\ \mu(d) &= 0 \quad \text{if any } \alpha_i > 1 \\ \mu(d) &= (-1)^k \quad \text{if } \alpha_1 = \alpha_2 = \cdots = \alpha_k = 1. \end{aligned}$$

The Möbius inversion formula for an arithmetic function is given by the following theorem [**1**, **2**, **4**]:

THEOREM 2. *Let $f(n)$ and $g(n)$ be arithmetic functions satisfying $f(n) = \sum_{d/n} g(d)$. Then $g(n) = \sum_{d/n} \mu(d) f(n/d)$, where $\mu(d)$ is the Möbius function.* ∎

THEOREM 3. *The number of irreducible monic polynomials of degree n over* GF(p), N_n, *is*

$$N_n = \frac{1}{n} \sum_{d|n} \mu(d) p^{n/d}. \tag{2}$$

Proof. Equation (2) is the Möbius inversion of equation (1). ∎

For example, N_{12} for $p=2$ is given by:

$$N_{12} = \frac{1}{12}(2^{12} - 2^6 - 2^4 + 2^2) = 335.$$

The surprising rate at which N_n increases for even small values of p is seen in the following table for $p=5$ and 7.

n \ p	5	7
2	10	21
3	40	112
4	150	588
5	624	3,360
6	2,580	19,544
7	11,160	117,648
8	48,750	720,300
9	217,000	4,483,696

By considering the representation of N_n given by equation (2) to be a p-ary number, it is easily shown that $N_n > 0$ for all positive n and p. This result provides another proof that there exists a monic irreducible polynomial of each degree n over GF(p) and, consequently, that there exists a finite field with p^n elements (obtained by extending GF(p) by the roots of the monic irreducible polynomial of degree n) for each prime p and integer n.

This work was supported by the US Atomic Energy Commission.

References

1. B. L. van der Waerden, Modern Algebra, Vol. 1, Ungar, New York, 1953, pp. 113–115.
2. Paul J. McCarthy, Algebraic Extensions of Fields, Blaisdell, Waltham, Mass., 1966, pp. 39–43 and p. 69.
3. R. D. Carmichael, Introduction to the Theory of Groups of Finite Order, Dover, New York, 1956 (Reprint), pp. 259–260.
4. J. B. Fraleigh, A First Course in Abstract Algebra, Addison-Wesley, Reading, Mass. 1968.

ON THE CONVOLUTION OF DISTRIBUTION FUNCTIONS

B. SCHWEIZER, University of Arizona

Let D be the set of distribution functions of one-dimensional random variables, i.e., functions defined on the real line R which are nondecreasing, left-continuous, and have inf 0 and sup 1. The binary operation of convolution is jointly continuous with respect to weak convergence on D. This fact is well known and a proof using characteristic functions is trivial. Such a proof is, however, non-elementary. The aim of this note is to provide an elementary proof and, in the process, an application of the Lévy metric as well. We shall need the following simple facts:

(1) The function $L:D\times D\to R$ defined by

$$L(F, G) = \inf_h\{F(x-h) - h \leqq G(x) \leqq F(x+h) + h, \quad \text{for all } x\}$$

is a metric, namely the *Lévy metric*, on D.

(2) Convergence in the metric space (D, L) is equivalent to weak convergence; that is to say, if $\{F_n\}$ is a sequence in D and F is also in D, then $L(F_n, F)\to 0$ if and only if F_n converges weakly to F (i.e., F_n converges pointwise to F at each continuity point of F).

(3) Convolution is a commutative operation on D. (This is usually derived from the corresponding property for characteristic functions but is easy to prove directly.)

THEOREM. *Convolution is jointly continuous on the metric space* (D, L).

Proof. Let $F, G\in D$ and $\epsilon>0$ be given. Suppose $F', G'\in D$ are such that $L(F, F')<\epsilon/2$ and $L(G, G')<\epsilon/2$, so that, for all $x\in R$,

$$F'\left(x-\frac{\epsilon}{2}\right)-\frac{\epsilon}{2} \leqq F(x) \leqq F'\left(x+\frac{\epsilon}{2}\right)+\frac{\epsilon}{2}$$

and

$$G'\left(x-\frac{\epsilon}{2}\right)-\frac{\epsilon}{2} \leqq G(x) \leqq G'\left(x+\frac{\epsilon}{2}\right)+\frac{\epsilon}{2}\cdot$$

Then

$$
\begin{aligned}
(F * G)(x) &= \int_{-\infty}^{\infty} F(x-y)dG(y) \leqq \int_{-\infty}^{\infty}\left[F'\left(x+\frac{\epsilon}{2}-y\right)+\frac{\epsilon}{2}\right]dG(y) \\
&= \int_{-\infty}^{\infty} F'\left(x+\frac{\epsilon}{2}-y\right)dG(y) + \frac{\epsilon}{2}\int_{-\infty}^{\infty} dG(y) \\
&= (F' * G)\left(x+\frac{\epsilon}{2}\right)+\frac{\epsilon}{2} = (G * F')\left(x+\frac{\epsilon}{2}\right)+\frac{\epsilon}{2} \\
&= \int_{-\infty}^{\infty} G\left(x+\frac{\epsilon}{2}-y\right)dF'(y)+\frac{\epsilon}{2} \\
&\leqq \int_{-\infty}^{\infty} G'(x+\epsilon-y)dF'(y)+\frac{\epsilon}{2}+\frac{\epsilon}{2} \\
&= (G' * F')(x+\epsilon)+\epsilon = (F' * G')(x+\epsilon)+\epsilon.
\end{aligned}
$$

Similarly, $(F' * G')(x-\epsilon)-\epsilon \leqq (F* G)(x)$, whence $L(F' * G', F * G) \leqq \epsilon$ and the theorem is proved.

MATHEMATICAL EDUCATION

EDITED BY J. G. HARVEY AND M. W. POWNALL

Material for this Department should be sent to either of the editors: J. G. Harvey, Department of Mathematics, University of Wisconsin, Madison, WI 53706; M. W. Pownall, Department of Mathematics, Colgate University, Hamilton, NY 13346.

THE QUESTION OF ACCREDITATION AND CERTIFICATION[1]

Introduction. Identification of persons who are really well qualified to teach mathematics in two- and four-year colleges is a perennial problem; and it is being aggravated by the growing number and diversity not only of schools to be staffed, but of graduate programs producing applicants. Similar problems exist in identifying persons who are qualified to work as mathematicians in industry or in other nonacademic positions. For these and other reasons, some consideration of the possibility of establishing a system of accreditation or certification in mathematics seems timely. The evidence [1] that we may be

[1] This is a digest of a *Report on Accreditation and Certification* prepared in 1969 by the CUPM Panel on College Teacher Preparation at the request of the MAA Board of Governors, and discussed by the Board on January 23, 1970. The Board appointed a committee of its members to survey mathematicians' views on this issue. The present digest, prepared by this committee, provides some background for questions that will be circulated through section officers within the next few months. The members of the committee are R. G. Bartle, Grace Bates, D. Bushaw, E. A. Cameron, D. T. Finkbeiner (chairman), J. Hashisaki, L. H. Lange, and F. L. Wolf.

PROBLEMS AND SOLUTIONS

All problems (both elementary and advanced) proposed for inclusion in this Department should be sent to E. P. Starke, 1000 Kensington Ave., Plainfield, NJ 07060. Proposers of problems are urged to enclose any solutions or information that will assist the editors. Ordinarily, problems in well-known textbooks and results in generally accessible sources are not appropriate for this Department. No solutions (except those accompanying proposals) should be sent to Professor Starke.

ELEMENTARY PROBLEMS

Solutions of Elementary Problems should be sent to Problems Group, Mathematics Department, University of Maine, Orono, ME 04473. To facilitate their consideration, solutions of Elementary Problems in this issue should be typed (with double spacing) and should be mailed before December 31, 1970. Contributors (in the United States) who desire acknowledgement of receipt of their solutions are asked to enclose self-addressed stamped postcards.

An asterisk means that no solution was submitted.

E 2247. *Proposed by N. S. Mendelsohn, University of Manitoba*

Let $a_1, a_2, \ldots, a_n$ be a set of relatively prime positive integers. Let $F(a_1, a_2, \ldots, a_n)$ represent the largest integer which cannot be represented in the form $c_1a_1+c_2a_2+\cdots+c_na_n$ where $c_1, c_2, \ldots, c_n$ are nonnegative integers. Prove the following:

(1) If $(a, b)=1$ and $a>0$, $b>0$ and c is a positive integer such that c is nonrepresentable in the form $Aa+Bb$, with A, B nonnegative integers, then $F(a, b, c)<F(a, b)$.

(2) If $(a, b)=1$, $a>0$, $a<b$ and such that $ta<b<(t+1)a$ where t is an integer, then $F(a, b, ab-(t+1)a-b)=ab-a-2b$.

E 2248. *Proposed by A. W. Walker, Toronto, Canada.*

If a, b, c, r, R are the side lengths, inradius, and circumradius of any plane triangle, then

$$\frac{1}{2rR} \leqq \frac{1}{3}\left(\frac{1}{a}+\frac{1}{b}+\frac{1}{c}\right)^2 \leqq \left(\frac{1}{a^2}+\frac{1}{b^2}+\frac{1}{c^2}\right) \leqq \frac{1}{4r^2}\cdot$$

E 2249.* *Proposed by A. L. Holshauer, Charlotte, N. C.*

Given any three triangles $\Delta_1, \Delta_2, \Delta_3$. Let H_1, H_2, H_3 be the orthocenters; G_1, G_2, G_3 the centroids; and O_1, O_2, O_3 the circumcenters. Prove that the centroids of the triangles $H_1H_2H_3$, $G_1G_2G_3$ and $O_1O_2O_3$ are collinear, and their circumcenters are likewise collinear.

E 2250. *Proposed by C. A. Kottman, Louisiana State University*

Prove that, if one is given any rectangular sheet of paper and a number $\epsilon > 0$, he may, by repeated foldings of the paper in half or in thirds (lengthwise or widthwise or both), arrive at a smaller rectangle with ratio r of length to width, satisfying $1-\epsilon \leqq r \leqq 1+\epsilon$.

E 2251. *Proposed by T. C. Brown, Simon Fraser University, Burnaby, British Columbia*

Consider a rectangular array of dots with an even number of rows and an even number of columns. Color the dots either red or blue in such a way that every row has the same number of red and blue dots, and likewise every column. Whenever two dots of the same color are adjacent in the same row or column, connect them with a line segment of that color. Show that the total number of blue segments equals the total number of red segments.

E 2252. *Proposed by Harry Lass, Jet Propulsion Laboratory, California Institute of Technology*

Given n urns numbered $1, \ldots, n$ and k objects, with $k \leqq n$. Suppose each of the objects is placed at random in one of the urns. Find for $r = 1, 2, \ldots, n$ the probability that the number of objects in the first r urns is less than or equal to r.

SOLUTIONS OF ELEMENTARY PROBLEMS

Editorial Note. Through an oversight the following names were omitted from the list of solvers of Problem E 2180 [1970, 406]: E. N. Fischman, R. Garfield, and Chung-kiu Wong.

An Equilateral Property of All Triangles

E 2139 [1968, 1114; 1969, 1066]. *Comment by A. W. Walker, Toronto, Canada.* The concluding statement in the published solution is incorrect. For a (30°, 60°, 90°) triangle ABC, O is the midpoint of AB, H coincides with C, and the apex of one of the equilateral triangles on base OH is therefore the point B, while the apex of the other is the reflection of O in AC and obviously does not coincide with an escribed center of triangle ABC.

A Product of Sixty-six Inversions

E 2140 [1968, 1114; 1969, 1067]. *Comment by A. W. Walker, Toronto, Canada.* The published solution is incomplete in that it omits an important case, namely when the mentioned orthocenter H coincides with the common point P of the three circles A, B, C. If H and P are distinct points, then (A_1, B_1, C_1) are circles with a common point P_1, as stated, but as $H \to P$, $P_1 \to \infty$; and when H and P coincide, then (A_1, B_1, C_1) are straight lines forming a triangle. The equivalence of the given relation and Thomsen's is then *direct*, and not inversive as stated in the solution. (For Thomsen's proof of his result, see *Math. Zeitschrift*, **34**, 1932, p. 690.)

A Functional Equation

E 2176 [1969, 554; 1970, 310]. *Proposed by R. S. Luthar, University of Wisconsin at Waukesha*

Find all continuous real functions f such that

$$f\left(\frac{x+y}{x-y}\right)=\frac{f(x)+f(y)}{f(x)-f(y)}. \tag{1}$$

Editorial Note. Part of the solution by L. E. Ward, Sr., was omitted in printing; the incompleteness of the published solution was pointed out by A. N. Aheart. A corrected solution follows.

Placing $y=0$ shows that $f(1)[f(x)-f(0)]=f(x)+f(0)$, whence we infer that $f(0)=0$ and $f(1)=1$, since f cannot be constant. Replacing y by $x-2$ and then x by $x-1$ and y by 1, we obtain

$$f(x-1)=\frac{f(x)+f(x-2)}{f(x)-f(x-2)}, \qquad f\left(\frac{x}{x-2}\right)=\frac{f(x-1)+1}{f(x-1)-1}. \tag{2}$$

It therefore follows that

$$f\left(\frac{x}{x-2}\right)=\frac{f(x)}{f(x-2)} \tag{3}$$

and

$$f(x)=\frac{f(x-1)+1}{f(x-1)-1}f(x-2). \tag{4}$$

By (3), $f(4)=f^2(2)$; this together with successive use of (4) yields

$$f(5)=\frac{f^2(2)+1}{[f(2)-1]^2}.$$

On the other hand,

$$f(5)=f\left(\frac{3+2}{3-2}\right)=\frac{f(3)+f(2)}{f(3)-f(2)}=\frac{f^2(2)+1}{1+2f(2)-f^2(2)}.$$

Hence $[f(2)-1]^2=1+2f(2)-f^2(2)$ or $f^2(2)=2f(2)$. But $f(2)\neq 0$ since, from (1), $f(2)=0$ would imply $f((x+2)/(x-2))\equiv 1$. Hence $f(2)=2$.

Now assume that $f(x)=x$ for $x=0, 1, 2, \ldots, n$ $(n\geqq 2)$. Then, with $x=n+1$, the first equation of (2) gives at once $f(n+1)=n+1$. It follows that $f(x)=x$ for all nonnegative integers.

Putting $y=cx$ yields

$$f\left(\frac{1+c}{1-c}\right)=\frac{f(x)+f(cx)}{f(x)-f(cx)}=\frac{1+f(c)}{1-f(c)},$$

from which $f(cx)=f(c)f(x)$. Now let $x=p/q$, a rational fraction, and $c=q$. Then $f(p)=f(p/q)f(q)$, showing that $f(x)=x$ for all positive rational numbers. By continuity it follows that $f(x)=x$ for all positive numbers.

Finally, placing $y=-x$ in (1) gives $f(0)=f(x)+f(-x)$, so that $f(-x)=-f(x)=-x$. We conclude that $f(x)\equiv x$.

A Special Property of 3

E 2190 [1969, 937]. *Proposed by Harry Pollard, Purdue University*

Show that if m and n are positive integers, the smaller of the numbers $\sqrt[n]{m}$ and $\sqrt[m]{n}$ cannot exceed $\sqrt[3]{3}$.

I. *Solution by Douglas Lind, Stanford University.* Let $f(x)=x^{1/x}$. Elementary calculus shows that $f(x)\to 0$ as $x\to 0$, $f(x)\to 1$ as $x\to\infty$, f is increasing in $[0, e]$, and f is decreasing in $[e, \infty)$. Hence $C=\sup(f(k):k=1, 2, \ldots)=\max(f(2),f(3))$. Since $3^2>2^3$, $f(3)>f(2)$, whence $C=3^{1/3}$. Then $f(m)\leqq 3^{1/3}$ for all positive integers m. If $n\geqq m$, then $m^{1/n}\leqq m^{1/m}\leqq 3^{1/3}$, whence $\min(m^{1/n}, n^{1/m})\leqq 3^{1/3}$.

II. *Solution by Charles Wexler, Arizona State University.* Suppose first that $m=n$, meaning that we must show that $\sqrt[n]{n}\leqq\sqrt[3]{3}$ or that $3^n\geqq n^3$. Proof by mathematical induction shows that this is true whenever $n\geqq 1$. ($3^n\geqq n^3\Rightarrow 3^{n+1}\geqq 3n^3=n^3+3n^2+3n+(n-3)n^2+(n^2-3)n$, which is greater than $(n+1)^3$ when $n\geqq 3$. The cases $n=1$, 2 are trivial.)

Next suppose that $1\leqq n<m$. Then $n^{1/m}\leqq n^{1/n}\leqq 3^{1/3}$.

Also solved by a hundred and seventeen other readers.

Another Variation of a Problem of Euler

E 2191 [1969, 938]. *Proposed by M. J. Zerger, Umpqua Community College, Roseburg, Oregon*

Find the solution set of $x^{(x+1)}=(x+1)^x$.

Solution by G. A. Heuer, Concordia College. There is only one solution, even if all different combinations of branches are considered in the case where $x<0$. This solution is an irrational number between 2 and e.

First assume $x>0$. A condition equivalent to the given one is that $f(x)=f(x+1)$, where $f(x)=(\log x)/x$. Inspection of $f'(x)$ shows that f is strictly increasing on $(0, e)$ and strictly decreasing on (e, ∞). It follows easily that there is a unique $x_0<e$ for which $f(x_0)=f(x_0+1)$. Since $f(2)<f(3)$, we know that $x_0>2$. The approximate value of x_0 is 2.2932.

For $x<0$, $(\log x)/x=[\log(x+1)]/(x+1)$ is still an equivalent condition. Now $\log x$ is not real, so $\log(x+1)$ is not; thus $x+1<0$ also. We consider all branches of the logarithm; thus we seek values of x, and integers m, n such that

$$\frac{\log|x+1| + i\pi(2m+1)}{x+1}=\frac{\log|x| + i\pi(2n+1)}{x}.$$

By equating imaginary parts we obtain $x=(2n+1)/(2m-2n)$, which requires x to be rational. By equating real parts we have

$$(\log|x+1|)/|x+1| = (\log|x|)/|x|,$$

since $x+1<0$. The only possible solution here would be $|x|=x_0+1$, $|x+1|=x_0$; i.e., $x=-x_0-1$. We eliminate this possibility by showing x_0 to be irrational.

Divide both members in the original equation by x^x to obtain

$$x = [(x+1)/x]^x,$$

and then $x^{1/x}=1+1/x$. Suppose $x_0=r/s$, where r and s are relatively prime integers. Then $(r/s)^{s/r}=1+s/r$;

$$r^s = s^s(1+s/r)^r = s^s(r+s)^r r^{-r}; \qquad r^{s+r} = s^s(r+s)^r.$$

This is impossible in view of $(r, s)=1$, unless $s=1$. But x_0 is not an integer, so it is not rational.

Also solved by Anders Bager (Denmark), Walter Bluger, G. C. Dodds, R. A. Gibbs, Michael Goldberg, Emil Grosswald, Thomas Hughes, P. G. Kirmser, Bill Knight, Patricia L. LaFratta, R. S. Lee, Charles Linnett, O. P. Lossers (Netherlands), D. E. Myers, C. S. Ogilvy, Stan Pauli, Corrado Quintiliani, Steven Russ, Simeon Reich (Israel), E. F. Schmeichel, E. M. Stone, Student Problems Group of St. Olaf College, Roger Weitzenkamp, Charles Wexler, and P. H. Young.

An Upper Bound

E 2192 [1969, 938]. *Proposed by R. S. Luthar, University of Wisconsin at Waukesha*

It is easy to show that

$$\sum_p \frac{1}{p^2} + \sum_p \frac{1}{p^3} + \sum_p \frac{1}{p^4} + \cdots < 1,$$

where the summation runs over all primes in each term of the above series. Determine a better upper bound.

Solution by O. P. Lossers, Technological University, Eindhoven, Netherlands.

$$\sum_{n=2}^{\infty}\sum_p \frac{1}{p^n} = \sum_p \sum_{n=2}^{\infty} \frac{1}{p^n} = \sum_p \frac{1}{p(p-1)} = \sum_p \left(\frac{1}{p-1} - \frac{1}{p}\right)$$
$$< \frac{1}{2} + \sum_{n=1}^{\infty}\left(\frac{1}{2n} - \frac{1}{2n+1}\right) = \frac{3}{2} - \log 2 < 0.8069.$$

Also solved by Anders Bager (Denmark), G. C. Dodds, R. B. Eggleton (Australia), T. E. Elsner, Leon Gerber, Michael Goldberg, Emil Grosswald, J. E. Hafstrom, C. T. Haskell, Dean Hickerson, Bill Knight, J. R. Kuttler, Douglas Lind, P. A. Lindstrom, D. C. B. Marsh, Arthur Marshall, Michael McCoy, C. B. A. Peck, R. S. R. Rao (India), Simeon Reich (Israel), Peter Ross, Steven Russ, E. F. Schmeichel, Ralph Schreiber, Peter Schroth (Germany), R. E. Shafer, J. S. Shipman, J. R. Smart, David Spear, S. E. Spielberg, Student Problems Group of St. Olaf

College, Hugo Sun, H. H. Thoyre, T. N. Tiwari, L. K. Tolman, E. W. Trost (Switzerland), Steve von Abele, Roger Weitzenkamp, Charles Wexler, J. E. Wilkins, Jr., and P. H. Young.

Note. Many contributors determined still better bounds. The citation for diligence goes to Wilkins who reports that the sum is 0.773156669049795128, correct to 18 decimal places.

A Weak Betweenness Relation

E 2193 [1969, 938]. *Proposed by D. J. Simanaitis, Case Western Reserve University*

For A, B and C, three points in the Euclidean plane, define B to be *weakly-between* A and C if and only if angle $ABC \geqq 120°$. Determine the minimal number of points required to insure the existence of at least one such weak-betweenness relation.

I. *Solution by Bill Knight, Student, University of Wyoming.* Let N be the minimum number of points needed. The regular polygons of sides 3, 4 and 5 show that N must be at least 6. It will be shown that $N=6$.

If A, B and C are collinear with B between A and C, then angle ABC will be greater than 120°, and B will be between A and C. If a point is in the interior of any triangle, it will be seen that that point is weakly-between some two of the vertices of the triangle.

Now assume that some six points are given in the plane in such a way as not to have a weakly-betweenness relation. The points are not collinear, so consider the convex hull, such that each of the six points is a vertex (since no three can be collinear), or is interior to the hull. The sum of the angles of any six-sided polygon is always 720°, so that at least one of the angles must be greater than or equal to 120°. It follows that one of the six points is in the interior. By connecting each pair of vertices of the hull, the interior is divided into triangles. Clearly, every point in the interior of the hull is also interior to one of these triangles and at least one weak-betweenness relation exists. Thus $N=6$.

II. *Solution by Jordi Dou, Barcelona, Spain.* El numero minimo es 6. Ya que entre los 5 vertices de un pentagono regular no hay weakly-between points (w.b.p.) y si A_i $(1 \leqq i \leqq 6)$ son 6 puntos, considérese la interseccion de los semiplanos $A_iA_j (i \neq j)$ que contienen los 6 puntos, si la interseccion no contiene ningan punto interior sera un exagono convexo y por lo menos un vertice sera w.b.p.; si contiene un punto A_i interior, A_i sera interior de un triangulo $A_jA_kA_l$ y sera w.b.p. de dos de sus vertices.

Also solved by Anders Bager (Denmark), J. C. Binz (Switzerland), Walter Bluger, W. G. Brady, E. P. Del Norte, Leon Gerber, Michael Goldberg, M. G. Greening (Australia), Ned Harrell, Dean Hickerson, D. C. Kay, J. R. Kuttler, Charles McCracken, R. A. Melter, Norman Miller, Robert Patenaude, Jürg Rätz (Switzerland), Simeon Reich (Israel), Kenneth Rosen, Steven Russ, Bill Sands, E. F. Schmeichel, Ralph Schreiber, G. F. Schumm, David Spear, H. H. Thoyre, Steve von Abele, Roger Weitzenkamp, Ruth Worth, Mark Yu, and the proposer.

Note. Melter notes that the solution follows from Exercise 2, p. 132 of Blumenthal, *Distance Geometry*. Rosen and Schmeichel found the problem to be substantially the same as Problem A-1 of the 1964 Putnam Competition (see this MONTHLY [1965, 735]).

A Four Square Problem

E 2194 [1969, 938]. *Proposed by Marion B. Smith, University of Wisconsin, Baraboo*

Do there exist nonzero integers a, b, c, d such that $a^2+b^2=c^2-d^2$ and $ab=cd$?

Solution by E. Rosenthall, McGill University. Using the idea of Fermat descent yields a negative answer. Clearly it suffices to determine that there are no solutions in positive integers. Suppose the system has a solution in positive integers and among all solutions choose one for which d is minimal. Then there are positive integers x, y, z, w with the conditions

$$(1) \qquad (x, z) = 1, \qquad (y, w) = 1$$

such that $a=kxy$, $b=kzw$, $c=kxw$, $d=kzy$, since all solutions of $ab=cd$ in positive integers are of this form. We also have $a^2+b^2=c^2-d^2$ and it follows that the positive integers x, y, z, w form a solution of

$$(x^2 + z^2)(y^2 + w^2) = 2x^2w^2,$$

where (1) then implies that only either z or y is even and also that $(x^2+z^2, x^2)=1$ and $(y^2+w^2, w^2)=1$. Then we have

$$x^2 + z^2 = w^2, \quad y^2 + w^2 = 2x^2, \qquad z \text{ even}$$

or

$$x^2 + z^2 = 2w^2, \quad y^2 + w^2 = x^2, \qquad y \text{ even.}$$

Considering the first system (the other system can be similarly considered) it follows from the first equation that there are relatively prime integers m, n of opposite parity with $m>n$ satisfying

$$x = m^2 - n^2, \quad z = 2mn, \quad w = m^2 + n^2$$

which, upon substitution into the second equation of the system yields

$$y^2 + (2mn)^2 = (m^2 - n^2)^2$$

and hence, since $(2mn, m^2-n^2)=1$, there are positive integers r, s with $r^2>s^2$ such that

$$y = r^2 - s^2, \quad mn = rs, \quad m^2 - n^2 = r^2 + s^2.$$

Then r, s, m, n is another solution of the given system in positive integers. But $d=k(2mn)(r^2-s^2)$ which implies $n<d$, and this contradicts the minimal property of d.

Also solved by Anders Bager (Denmark), Walter Bluger, W. J. Blundon, W. F. Fox, Michael Goldberg, M. G. Greening (Australia), Thomas Hughes, J. A. H. Hunter, O. P. Lossers (Netherlands), D. C. B. Marsh, W. W. Meyer, Simeon Reich (Israel), K. A. Ribet, E. F. Schmeichel, Ralph Schreiber, David Spear, Allen Stenger, E. W. Trost (Switzerland), T. N. Tiwari, Steve von Abele, Charles Wexler, and the proposer.

Note. Most solvers note that the hypothesis implies $(ad+bc)^2=c^4-a^4$ and that this Diophantine equation is well known as having no solution.

A Divisibility Problem in Polynomial Rings

E 2195 [1969, 938]. *Proposed by T. J. Bruggeman, Xavier University, Cincinnati, Ohio*

Which polynomials of the form $\sum_{i=1}^{n} x^{a_i}$ are divisible by $\sum_{i=1}^{m} x^{i-1}$? That is, find necessary and sufficient conditions for a_i, $i=1, 2, \cdots, n$.

Solution by Robert Gilmer, Florida State University. We assume that $f(x)=\sum_{i=1}^{m} x^{i-1}$ and $g(x)=\sum_{i=1}^{n} x^{a_i}$ are elements of $R[x]$, where R is a commutative ring with unity e. We write a_i as $q_i m+r_i$, where $0 \leqq r_i < m$. Then, since $x^m \equiv e$ (mod $f(x)$), we have $g(x) \equiv \sum_{i=1}^{n} x^{r_i}$ (mod $f(x)$). For any j between 0 and $m-1$, we let n_j be the number of a_i's which are congruent to j modulo m. Then $\sum_{i=1}^{n} x^{r_i} =\sum_{j=0}^{m-1} n_j e x^j = h(x)$; since $f(x)$ is monic, it is clear that $f(x)$ divides $h(x)$ if and only if $n_0 e = n_1 e = \cdots = n_{m-1} e$. Thus, $f(x)$ divides $g(x)$ if and only if n_i and n_j are congruent modulo the characteristic of R for any i and j between 0 and $m-1$. One interesting corollary of this result is that m must divide n if $f(x)$ divides $g(x)$.

Also solved by Daniel Frohardt, Leon Gerber, Michael Goldberg, C. S. Ogilvy, Simeon Reich (Israel), E. F. Schmeichel, Ralph Schreiber, Sid Spital, Steve von Abele, and the proposer.

Note. Problem E 2000 [1968, 908] is a special case.

Multiplicative Function and its Dirichlet Inverse

E 2196 [1969, 1062]. *Proposed by R. Sivaramakrishnan, Trichur, India*

Given a multiplicative arithmetic function $f(n)$, let its Dirichlet inverse be $f^{-1}(n)$. By definition, $f^{-1}(n)$ satisfies the equation

$$\sum_{d|n} f(d)f^{-1}(n/d) = \begin{cases} 1 & (n=1) \\ 0 & (n>1). \end{cases}$$

Prove that $\sum_{d|n} f(d)f^{-1}(n/d)d = f(n)\phi(n)$ if and only if $f(n)$ is completely multiplicative. $\phi(n)$ is the well-known Euler function.

Solution by Dan Marcus, Harvard University. Rather than assume that f is multiplicative, we assume only that $f(1) \neq 0$; this guarantees the existence of f^{-1}.

If A is the set of all arithmetic functions (complex-valued if we wish) which do not vanish at 1, then it is well known that A is a commutative group under the Dirichlet product

$$(f*g)(n) = \sum_{d|n} f(d)g(n/d),$$

and that A is a commutative monoid under pointwise multiplication $(fg)(n) = f(n)g(n)$. Obviously the pointwise identity is the constant function 1, and the Dirichlet identity is $(f*f^{-1})(n) = \delta_{1n}$ (Kronecker delta). If we let $I(n)=n$, then

the problem asks us to show that $(fI)*f^{-1}=f\phi$ if and only if f is completely multiplicative; this is evidently equivalent to showing that $(f\phi)*f=fI$ if and only if f is completely multiplicative.

Suppose that f is completely multiplicative. Then pointwise multiplication by f distributes over the Dirichlet product [J. Lambek, *Arithmetical functions and distributivity*, this MONTHLY 73 (1966) 969–973]. Since $\phi*1=I$ (a well-known number theoretic result), we have that

$$fI = f(\phi * 1) = (f\phi) * (f1) = (f\phi) * f,$$

as was to be shown.

Conversely, if $fI=(f\phi)*f$, then in particular $f(1)=(f(1))^2$ so that $f(1)=1$. For $n>1$, we have that

$$(n - 1 - \phi(n))f(n) = \sum_{d\mid n}{}' f(d)\phi(d)f(n/d), \tag{1}$$

where the prime on the sum indicates that we sum over all divisors of n except n and 1. Now define g by defining $g(p)=f(p)$ if $p=1$ or p is prime, and extending g to all n so as to be completely multiplicative. By the first part of the problem

$$(n - 1 - \phi(n))g(n) = \sum_{d\mid n}{}' g(d)\phi(d)g(n/d), \tag{2}$$

holding for all $n>1$. Equations (1) and (2) now imply that if n is composite and if $f(d)=g(d)$ for all $d<n$, then $f(n)=g(n)$, since $\phi(n)\neq n-1$ if n is composite. Thus $f=g$ by complete induction, and f is completely multiplicative.

Also solved by B. Averbach, Trygve Breiteig (Norway), L. Carlitz, N. J. Fine, M. G. Greening (Australia), Emil Grosswald, Douglas Holdridge, Wells Johnson, R. Sita Ramachendra Rao (India), David Rearick, Simeon Reich (Israel), E. F. Schmeichel, E. A. Smith, E. W. Trost (Switzerland), and the proposer.

Editorial Note. Most solvers noted that if f is completely multiplicative, then $f^{-1}=f\mu$, where μ is the Möbius function. Thus

$$\begin{aligned}\sum_{d\mid n} df(d)f^{-1}(n/d) &= \sum_{d\mid n} df(d)f(n/d)\mu(n/d)\\ &= f(n)\sum_{d\mid n} d\mu(n/d) = f(n)\phi(n).\end{aligned}$$

(The fact that $I*\mu=\phi$ is well known; M. W. Potter supplied the reference, Cashwell and Everett, *The ring of number-theoretic functions*, Pacific J. Math., 9 (1959) 975–985). The converse was generally shown by assuming that f was multiplicative and then proving by induction that $f(p^k)=(f(p))^k$ for p prime. Carlitz and Rearick noted that f need not be assumed to be multiplicative to show this.

Lambek (*ibid.*) showed that f is completely multiplicative if and only if f distributes over *all* Dirichlet products: $f(g*h)=fg*fh$ for all g, h. The present problem asserts essentially that f is completely multiplicative if and only if $f(\phi*1)=f\phi*f1$, i.e., that f distributes over the *particular* Dirichlet product $\phi*1=I$.

The Functional Equation $F(x^m)=[F(x)]^n$

E 2197 [1969, 1063]. *Proposed by M. S. Klamkin, Ford Scientific Laboratory, and D. J. Newman, Yeshiva University*

Solve the functional equation $F(x^m)=[F(x)]^n$.

Solution by the proposers. Let $\log F(x)=G(x)(\log x)^\alpha$, where $\alpha=(\log n)/(\log m)$. Then $G(x^m)=G(x)$. A general solution for $G(x)$ is $G(x)=H(\log\log x)$, where H is periodic with period $\log m$. Thus

$$F(x) = \exp[H(\log\log x)\cdot(\log x)^\alpha].$$

It is to be noted that the problem was deliberately incompletely formulated in that no class of functions F and no domain of x were specified, nor the constants m, n. If $F(x)$ is to be real, then it is assumed that $x>1$. In the above it is also assumed that m, $n>0$ and $m\neq 1$. (The case $m=1$ is easily handled. If $mn=0$, the equation may be solved by inspection. The above solution is also valid when m, n are not both positive, provided α can be chosen so that $m^\alpha=n$.)

Also solved by Michael Amling, E. P. Del Norte, G. A. Edgar, Leon Gerber, G. A. Heuer, and E. W. Trost (Switzerland).

Sidney Penner had earlier posed the special case $m=n=2$.

ADVANCED PROBLEMS

All solutions of Advanced Problems should be sent to J. Barlaz, Rutgers—The State University, New Brunswick, N.J. 08903. Solutions of Advanced Problems in this issue should be typed (with double spacing) on separate, signed sheets and should be mailed before December 31, 1970. Contributors (in the United States) who desire acknowledgment of receipt of their solutions are asked to enclose self-addressed stamped postcards.

An asterisk means that no solution was submitted.

5736 [1970, 532]. **Correction.** *Proposed by M. S. Klamkin, Ford Scientific Laboratory*

Solve the nonlinear difference equation of rth order

$$D_n = a_1D_{n-1}^{m+1} + a_2D_{n-1}^{m}D_{n-2}^{m+1} + \cdots + a_rD_{n-1}^{m}D_{n-2}^{m}\cdots D_{n-r+1}^{m}D_{n-r}^{m+1},$$

(m, r, a_i, constants).

5746. *Proposed by Leonard Carlitz, Duke University*

Let $GF(2^n)$ denote the finite field or order 2^n. For $a\in GF(2^n)$ put

$$e(a) = (-1)^{t(a)},\quad t(a) = a + a^2 + a^{2^2} + \cdots + a^{2^{n-1}},$$

$$S(a) = \sum_{x,y,z} e\left\{x+y+z+\frac{a}{yz+zx+xy}\right\},$$

where the summation is over all $x, y, z \in GF(2^n)$ such that $yx+zx+xy \neq 0$. Show that

$$S(a) = (-1)^n 2^n \sum_{x \neq 0} e(x + ax'), \qquad (xx' = 1).$$

5747. *Proposed by H. M. Edgar and Martin Billik, San Jose State College*

Let $1 < n_1 < n_2 < \cdots < n_k$ be integers. Let the integers b_i satisfy $0 \leqq b_i \leqq n_i - 1$ for every value of i with $1 \leqq i \leqq k$ and assume that (n_i, n_j) is not a divisor of $(b_i - b_j)$ for all $i \neq j$, $1 \leqq i, j \leqq k$. Prove that there must exist an integer x satisfying $x \not\equiv b_i \pmod{n_i}$ for all i with $1 \leqq i \leqq k$.

5748. *Proposed by R. L. Graham, Bell Telephone Laboratories.*

Let $0 < a_n < a_{n+1} + a_{n^2}$, $n \geqq 1$. Show that $\sum_{n=1}^{\infty} a_n$ diverges.

5749.* *Proposed by R. L. Graham, Bell Telephone Laboratories*

Let $0 < a_1 < \ldots < a_n$ be integers. If (a, b) denotes the greatest common divisor of a and b, show that

$$\frac{a_i}{(a_i, a_j)} \geqq n$$

for some i and j.

5750. *Proposed by John Horvath, University of Maryland*

Let f be a decreasing, bounded function, defined on the interval $0 \leqq x \leqq 1$ of the real line. Prove that there exists a sequence (f_n) of continuous, decreasing functions having the same bounds as f, which converges almost everywhere to f.

5751. *Proposed by Marvin Marcus, University of California at Santa Barbara*

Let S_m denote the symmetric group of degree m. For $\sigma \in S_m$ let $c(\sigma)$ be the number of cycles (including cycles of length 1) in the disjoint cycle decomposition of σ. Prove that for $1 \leqq m \leqq n$,

$$\binom{n}{m} = \frac{1}{m!} \sum_{\sigma \in S_m} (\operatorname{sgn} \sigma) n^{c(\sigma)}. \tag{1}$$

Also prove that for any positive integers m and n,

$$\binom{n + m - 1}{m} = \frac{1}{m!} \sum_{\sigma \in S_m} n^{c(\sigma)}. \tag{2}$$

5752.* *Proposed by Kesiraju Satyanarayana, Rajahmundry, India*

If

$$\sigma_{m,r} = \sum_{p=0}^{m} (-1)^p \binom{2m+1}{p}^r \sum_{x=p+1}^{2m+1-p} \frac{1}{x}$$

is expressed as a fraction in lowest terms, show that

(1) $\sigma_{m,1}=1/(2m+1)$.

(2) For $m=1$ to 7, the following results hold:

(i) $\sigma_{m,2}$, $\sigma_{m,3}$ have the sign of $(-1)^m$.

(ii) The numerator of $|\sigma_{m,2}|$ is a power of 2.

(iii) The numerator of $|\sigma_{m,3}|/(2m+1)!$ is the product of some successive prime numbers immediately following $2m+1$.

(3) Do the results in (2) hold in general?

SOLUTIONS OF ADVANCED PROBLEMS

On *ABA*-Groups

5673 [1969, 565; 834]. *Proposed by Stanley E. Payne, Miami University, Ohio*

Let G be a group with subgroups A and B such that $(A:A\cap B)=(B:A\cap B)=k\geqq 3$, and such that there is an $n<\infty$ and an $x\in G$ with Ax^i, Bx^i, $1\leqq i\leqq n$, being all the distinct left cosets of A and B in G. Then prove that $ABA\cap BAB\neq AB+BA$ if $AB+BA=A+B$.

Solution by the proposer. Consider the coset geometry $\pi(G, A, B)$ in which points are all left cosets Ax and lines are all left cosets Bx, $x\in G$. Then Ax is incident with By if and only if $Ax\cap By\neq\emptyset$. A chain of incident points and lines starting with A must have the form A, Ba_1, Ab_1a_1, $Ba_2b_1a_1$, $\cdots$ for which the chain is irreducible if and only if each $a_i\notin B$ and each $b_i\notin A$. Here $A:A\cap B=k\geqq 3$ implies A is incident with at least three distinct lines B, Bx^i, Bx^j. Then A, B, Ax^{-i}, Bx^{-i+j}, Ax^{-i+j}, Bx^j, A is a closed irreducible chain, whence $A=Ab_3a_3b_2a_2b_1a_1$ for $a_i\in A\backslash B$, $b_i\in B\backslash A$, which is essentially what we needed to show. (For reference see Higman and McLaughlin, *Geometric ABA-groups*, Ill. J. Math., 5(1961) 382–397.)

Complete Metric and the Real Line

5676 [1969, 566]. *Proposed by Albert Wilansky, Lehigh University*

Let $X=\{0\}\cup(1, \infty)$ with the ordinary topology of the real line. Suppose that a complete metric is given for this topology. Must $(1, \infty)$ contain a closest point to $\{0\}$?

Solution by W. O. Alltop, China Lake, California. We answer the question negatively. Clearly X is homeomorphic to $y=\{P\}\cup R$, where R is the real line with the usual topology, and P is a point isolated from R, i.e., $\{P\}$ is an open set. For x, $y\in R$ let

$$d(x, y) = |x-y|/(1+|x-y|).$$

This well-known distortion of the usual metric leaves (R, d) homeomorphic to $(R, |\,|)$. Moreover, the Cauchy sequences of (R, d) are precisely those of

$(R, |\,|)$. Thus, (R, d) is complete. Now for $x \in R$, let

$$d(P, x) = (2 + |x|)/(1 + |x|).$$

This distance from P to R is 1, but this distance is not achieved for any point of R. It remains only to show that (Y, d) is a metric space, i.e., that the triangle inequality holds for d. For $x, y \in R$ we have

$$d(P, x) + d(P, y) > 2 > 1 > d(x, y).$$

Also

$$\begin{aligned} |d(P, y) - d(P, x)| &= \frac{||x| - |y||}{1 + |x| + |y| + |x||y|} \\ &\leqq \frac{|x - y|}{1 + |x - y|} = d(x, y). \end{aligned}$$

Thus, $d(P, y) \leqq d(P, x) + d(x, y)$, so d is a metric for Y.

Also solved by Linda W. Brinn, J. M. Cibulskis, D. E. Cooper, E. P. Del Norte, L. J. Dickson, Michael Golomb, D. A. Hejhal, Ellen Hertzmark & W. C. Waterhouse, D. W. Hodwin, H. C. Kranzer, M. D. Mavinkurve (India), W. G. McArthur, Torquist Memp, J. T. Rosenbaum, D. A. Zave and Martin Ziegler (Clausthal).

The Spectral Radius of a Matrix

5677 [1969, 701]. *Proposed by B. W. Levinger, Case Western Reserve University*

Let $A = (a_{ij})$ be a real $n \times n$ matrix with $a_{ij} \geqq 0$, $1 \leqq i, j \leqq n$. Prove that $r(A) \leqq r[\frac{1}{2}(A + A^T)]$, where $r(C)$ denotes the spectral radius of a matrix C.

I. *Solution by Emeric Deutsch, Polytechnic Institute of Brooklyn.* By Frobenius' theorem (see e.g. P. Lancaster, *Theory of Matrices*, Academic Press, New York-London, 1969, p. 288), $r(A)$ is an eigenvalue of A. Now by Bendixson's theorem (see e.g. A. S. Householder, *The Theory of Matrices in Numerical Analysis*, Blaisdell, New York-Toronto-London, 1964, p. 69), we have $\beta_n \leqq r(A) \leqq \beta_1$, where $\beta_1 \geqq \beta_2 \geqq \cdots \geqq \beta_n$ are the eigenvalues of $\frac{1}{2}(A + A^T)$. Hence $r(A) \leqq r[\frac{1}{2}(A + A^T)]$.

II. *Solution by Bertram Walsh, University of California at Los Angeles.* The result of the problem is generalized by the following two propositions:

PROPOSITION I. Let H be a (complex) Hilbert space, $A \in \mathcal{L}(H)$, and let $\mathrm{Re}A = \frac{1}{2}(A + A^*)$. If $\lambda \in \sigma(A)$, then $\mathrm{Re}\lambda \leqq \|\mathrm{Re}A\| = r(\mathrm{Re}A)$.

PROPOSITION II. Let E be an ordered (real) Banach space, and let K denote its cone of nonnegative elements. Suppose that $E = K - K$ and that K is normal in the sense that the topology of E is generated by a norm which is an increasing

function on K. If $A\in\mathcal{L}(E)$ is nonnegative (i.e., if $A[K]\subseteq K$), then $r(A)\in\sigma(A)$, where the spectrum is taken in the complexification of E. [This may be found in H. Schaefer, *Topological Vector Spaces*, 1966, Appendix, Thm. 2.2, p. 263. The matrix version goes back to Frobenius (S.-B. Preuss. Akad. Wiss., Berlin 1908, 471–476; 1909, 514–518; 1912, 456–477.)]

To put these propositions together to solve the original problem, one need only give $\boldsymbol{R}^n$ the Euclidean norm and the coordinatewise order; its complexification is the Hilbert space $\boldsymbol{C}^n$, and one may set $\lambda=r(A)$ in Proposition I by virtue of Proposition II. More generally, one may take any nonnegative operator A on $L^2_R(\mu)$, where (X, S, μ) is a given measure space, and derive the inequality $r(A)\leqq r(\mathrm{Re}A)$; the complexification of $L^2_R(\mu)$ in this setting is of course $L^2_C(\mu)$.

Also solved by G. P. Barker, J. R. Kuttler, Henryk Minc, R. K. Mueller, R. C. Thompson, and the proposer.

Note. Barker notes that the result is generalized by replacing the condition $a_{ij}\geqq 0$ by its consequence that $r(A)$ is an eigen-value of A.

Converegnce Set for $\sin(\lambda_n x+\epsilon_n)$

5679 [1969, 701]. *Proposed by H. Kestelman, University College, London, England*

$\{\lambda_n\}$ and $\{\epsilon_n\}$ are real number sequences with $\lim_{n\to\infty}\lambda_n=\infty$. Prove that the set of x for which the sequence $\{\sin(\lambda_n x+\epsilon_n)\}$ converges has Lebesgue measure zero but may have cardinal c.

Solution by H. E. Bray, Rice Institute. We may express ϵ_n as $2\pi k_n+\delta_n$, k_n an integer, $0\leqq\delta_n<2\pi$. Thus $\sin(\lambda_n x+\epsilon_n)=\sin(\lambda_n x+\delta_n)$, where the sequence $\{\delta_n\}$, being bounded, contains a subsequence converging to a number δ. Hence if $\sin(\lambda_n x+\epsilon_n)$ converges so does $\sin(\lambda_n' x+\delta_n')$, where $\{\lambda_n'\}$, $\{\delta_n'\}$ are corresponding subsequences of $\{\lambda_n\}$, $\{\delta_n\}$ and $\lim\delta_n'=\delta$. We may now drop primes and suppose that $\sin(\lambda_n x+\delta_n)$ converges on a set E which is of course measurable since the lim sup and lim inf of a sequence of measurable functions are measurable.

$$\begin{aligned}\sin(\lambda_n x+\delta_n) &= \sin\lambda_n x\cos\delta_n+\cos\lambda_n x\sin\delta_n\\ &= \sin\lambda_n x\cos\delta+\cos\lambda_n x\sin\delta+O(1).\end{aligned}$$

Hence $\int\sin(\lambda_n x+\delta_n)$ converges to zero on every measurable set E, by a well known property of the Lebesgue integral, and therefore $\sin(\lambda_n x+\delta_n)$ cannot converge to a positive (negative) function on a set of positive measure.

It remains to prove that $\sin(\lambda_n x+\epsilon_n)$ cannot converge to zero on a set E of positive measure. Supposing on the contrary that $\sin^2(\lambda_n x+\epsilon_n)$ converges to zero on E and writing:

$$2\sin^2(\lambda_n x+\epsilon_n)=1-\cos(2\lambda_n x+2\epsilon_n),$$

it follows that

$$\cos(2\lambda_n x+2\epsilon_n)=\sin(2\lambda_n x+2\epsilon_n-\tfrac{1}{2}\pi)$$

converges to unity on E. But this is impossible if $|E|>0$ as we have proved. Hence $\sin(\lambda_n x+\epsilon_n)$ cannot converge on a set of positive measure.

If now we consider all values of x representable in the binary scale in the non-terminating form:

$$x = .a_1 0 a_2 00 a_3 000 a_4 0 \cdots, \quad a_j = 0 \text{ or } 1, \quad \sum a_j = \infty,$$

where a_n is followed by n zeros preceding a_{n+1} and occupies the (p_n)th place, where $p_n = n(n+1)/2$, it follows that if $\lambda_n = 2^{p_n} \cdot 2\pi$ then

$$\lambda_n x = 2k_n\pi + 2\pi(.000 \cdots a_{n+1}00 \cdots a_{n+2}00 \cdots)$$

where k_n is an integer determined by x and n.

And since the second term is less than $2\pi/2^n$ it follows that $\sin \lambda_n x$ converges to zero on a set of values of x which are in one-to-one correspondence with the set of non-terminating numbers of the form:

$$y = .a_1 a_2 a_3 \cdots, \qquad \sum a_n = \infty.$$

Each point of the interval $0<y\leqq 1$ is representable in one and only one way by a number of this set. Hence $\sin\lambda_n x$ converges to zero on a set of measure zero and of cardinal number c.

Also solved by G. J. Foschini, D. A. Hejhal, R. K. Meany, P. van der Steen (Netherlands), and the proposer.

Ideals in a Dedekind Domain

5680 [1969, 701]. *Proposed by Kamlesh Wasan, University of Delhi, India*

Consider a Noetherian integrally closed domain R in which each semi-primary ideal (i.e., an ideal whose radical is a prime ideal) is irreducible. Prove R is a Dedekind domain.

I. *Solution by James R. Smith, Appalachian State University.* By a theorem in Zariski and Samuel, *Commutative Algebra*, Vol. I, p. 275, all that is needed is to prove that every proper prime ideal of R is maximal. So, let P be a prime ideal of R, M a maximal ideal which contains P. Then $P\cap M^n$ is a semi-primary ideal for all n, for if $x\in P$, $x\in M$, so $x^n \in P\cap M^n$. Thus $P\cap M^n = P$ or $P\cap M^n = M^n$. If for some n, $M^n = P\cap M^n$, then $P=M$ since P is prime, and we are finished. If for every n, $P\cap M^n = P$, then $P\subset \bigcap_{n\in N} M^n = 0$ (Zariski and Samuel, p. 216).

II. *Solution by Robert Gilmer, Florida State University.* The following more general result is true:

If R is a Noetherian commutative ring with identity, and if for each maximal ideal M of R, M^2 is irreducible, then each ideal of R is a finite product of prime ideals.

Proof. K. Asano (Journal of the Mathematical Society of Japan, 3(1951), pp. 82–90, Satz 5) has proved that each ideal of a commutative ring S with identity is a finite product of prime ideals if and only if S is Noetherian and

there are no ideals properly between M and M^2 for each maximal ideal M of S. Hence we show that the given hypothesis on R implies that there are no ideals properly between M and M^2 for each maximal ideal M of R.

M/M^2 is canonically a vector space over the field R/M under the scalar multiplication $(r+M)(m+M^2)=rm+M^2$. Further, the mapping $A\to A/M^2$ is a one-to-one correspondence between the set of ideals of R lying between M and M^2 and the set of subspaces of M/M^2, and this correspondence preserves intersection. Since M^2 is an irreducible ideal of R, the zero subspace of M/M^2 is irreducible; this implies, however, that M/M^2 has dimension at most one and, consequently, there are no ideals of R properly between M and M^2.

Also solved by M. L. Laplaza, Ka Menehune, and the proposer.

Zeros of Hermite-type Polynomials

5681 [1969, 702]. *Proposed by C. L. Sabharwal, St. Louis University*

Let $H_n(x)$ be a monic Hermite polynomial of degree n, given by

$$H_n(x) = (-1)^n e^{x^2/2}\frac{d^n}{dx^n}(e^{-x^2/2}),$$

and let q and m be nonnegative integers with $q+m$ even. Show that

$$P(x) = \sum_{\gamma=0}^{\min(q,m)}(-1)^\gamma\binom{q}{\gamma}\frac{m!}{(m-\gamma)!}x^{m-\gamma}H_{q-\gamma}(x),$$

has $q+m$ distinct nonzero real roots if $q>m$, and $2q$ distinct nonzero real roots if $q\leqq m$.

Solution by O. P. Lossers, Technological University, Eindhoven, The Netherlands. Substitute the Hermite polynomial in $P(x)$; then by means of Leibniz' formula one has

$$\begin{aligned}(1)\qquad P(x) &= (-1)^q e^{x^2/2}\sum_{\gamma=0}^{q}\binom{q}{\gamma}\frac{d^\gamma}{dx^\gamma}(x^m)\frac{d^{q-\gamma}}{dx^{q-\gamma}}(e^{-x^2/2})\\ &= (-1)^q e^{x^2/2}\frac{d^q}{dx^q}(x^m e^{-x^2/2}).\end{aligned}$$

We distinguish the following cases:

(i) $m=0$. Then one has

$$P(x) = (-1)^q e^{x^2/2}\frac{d^q}{dx^q}(e^{-x^2/2}) = H_q(x).$$

It is well known that the Hermite polynomial $H_q(x)$ has q distinct real roots. If q is even all roots are nonzero; if q is odd one root is zero and the remaining $q-1$ roots are different from zero.

(ii) $q=0$. Then one has $P(x)=x^m$, i.e., $P(x)$ only has a zero root of multiplicity m.

(iii) $m>0$, $0<q\leqq m$. In view of (1), $P(x)$ has a zero root of multiplicity $m-q$. Because of $x^m e^{-x^2/2}\to 0$ when $x\to\pm\infty$, it follows from Rolle's theorem, using induction on q, that $P(x)$ has $2q$ distinct nonzero real roots.

(iv) $m>0$, $q>m$. According to Rolle's theorem $P(x)$ has $q+m$ distinct real roots. If $q+m$ is even all roots are nonzero; if $q+m$ is odd one root is zero and the remaining $q+m-1$ roots are different from zero.

Also solved by the proposer.

Convergence Domains of Power and Dirichlet Series

5682 [1969, 702]. *Proposed by Robert Spira, Michigan State University*

What is the relation of the radius of convergence of $\sum a_n z^n$ and the abscissa of convergence of $\sum a_n n^{-z}$?

Solution by A. C. Hindmarsh, Livermore, California. The relation between the radius of convergence R of $\sum a_n z^n$ and the abscissa of convergence A of $\sum a_n n^{-z}$ is as follows:

(a). If $R>1$, then $\limsup |a_n|^{1/n}<1$. Thus, for some $a<1$ we have $|a_n|<a^n$ for all large n. Hence $\sum a_n n^{-z}$ is dominated by $\sum a^n n^{-x}$, with $x=\mathrm{Re}(z)$, and this converges for all x. Therefore $A=-\infty$.

(b). If $R<1$, then $\limsup |a_n|^{1/n}>1$, and for some $a>1$, we have $|a_n|>a^n$ for infinitely many n. For such n, $|a_n n^{-z}|>a^n n^{-x}$, which tends to ∞ as $n\to\infty$. Therefore $\sum a_n n^{-z}$ diverges for all z, or $A=+\infty$.

(c). If $R=1$, we may have A finite, as in the case $a_n=n^A$, or $A=-\infty$, as when $a_n=n^{-\log n}$, or $A=+\infty$, as when $a_n=e^{\sqrt{n}}$.

Also solved by D. A. Hejhal, M. D. Mavinkurve (India), and the proposer.

Note. Mavinkurve calls attention to the following which appears in M. A. Evgratov, *Collection of Problems in the Theory of Analytic Functions*, Moscow, 1969, 12.03. Let $\{\lambda_n\}$ be an increasing sequence of positive numbers and let $\limsup_{n\to\infty}(n/\lambda_n)=\alpha>0$, and $\limsup |a_n|^{1/n}=\rho>0$. Show that the sum of the series $\sum_1^\infty a_n e^{-\lambda_n z}$ is regular in the halfplane $\mathrm{Re}\, z>\alpha\log\rho$.

"Packing" the Unit Circle

5684 [1969, 835]. *Proposed by Steven Minsker, Massachusetts Institute of Technology.*

Let $U=\{z:|z|<1\}$ in the complex plane. We pick a sequence of open disks U_n of radius r_n such that the following three conditions are satisfied:

$$\text{(a) } \overline{U}_n\subset U, \qquad \text{(b) } \overline{U}_i\cap\overline{U}_j=\varnothing, \qquad \text{(c) } \sum_{n=1}^{\infty} r_n<\infty,$$

$n=1, 2, 3, \cdots, i\neq j$. Let $X=U-\bigcup_{n=1}^\infty U_n$. Prove that X has positive Lebesgue measure.

Editorial Note. As pointed out by several readers, this problem has been solved and generalized. See number 5528 [1968, 913].

Also solved by R. O. Davies (England), D. A. Hejhal, R. A. Horn, Douglas Lind, J. B. Linder, O. P. Lossers (Netherlands), Simeon Reich (Israel), and the proposer.

The Arithmetic-Geometric Mean Inequality

5685 [1969, 835]. *Proposed by D. E. Daykin, University of Malaya, Kuala Lumpur*

Let m, n be positive integers and $a_1, a_2, \cdots, a_n$ be positive reals. For $i=1, 2, 3, \cdots$, put $a_{n+i}=a_i$ and $b_i=a_{i+1}+a_{i+2}+\cdots+a_{i+m}$. Then show that $m^n a_1 a_2 \cdots a_n < b_1 b_2 \cdots b_n$, except if all the a_i are equal.

Solution by E. F. Schmeichel, College of Wooster. The desired inequality is equivalent to $\prod_{i=1}^n (b_i/m) > a_1 a_2 \cdots a_n$. Noting that

$$\frac{b_i}{m} = \frac{1}{m}\sum_{k=1}^m a_{i+k} \geqq \left(\prod_{k=1}^m a_{i+k}\right)^{1/m},$$

with equality only if all the a_{i+k} are equal, we have

$$\prod_{i=1}^n \left(\frac{b_i}{m}\right) > \prod_{i=1}^n \left(\prod_{k=1}^m a_{i+k}\right)^{1/m} = \left(\prod_{i=1}^n a_i^m\right)^{1/m} = a_1 a_2 \cdots a_n,$$

unless all the a_i are equal. (It should be noted that for $m=1$ and any set $\{a_i\}_{i=1}^n$, the inequality must be replaced by an equality.)

Also solved by J. C. Binz (Switzerland), John Bessman, M. T. Bird, L. Carlitz, C. S. Karuppan Chetty (India), R. O. Davies (England), N. J. Fine, David Friedman, M. G. Greening (Australia), Robert Heller, G. A. Heuer, M. S. Klamkin, R. A. Kopas, J. R. Kuttler, Gérard Letac (France), O. P. Lossers (Netherlands), Andrej Mąkowski (Poland), R. H. C. Newton (Wales), P. J. Owens (England), Simeon Reich (Israel), T. Tamura (Japan), A. C. Williams, Aleksandras Zujus, and the proposer.

Klamkin uses an extension of Hölder's inequality

$$\sum_{i=1}^m \prod_{j=1}^n x_{ij}^{1/p_j} = \prod_{j=1}^n \left(\sum_{i=1}^m x_{ij}\right)^{1/p_j}$$

to demonstrate the problem. (See Beckenbach and Bellman, *Inequalities*, p. 20.)

Archimedean Ordered Subfields of $Q(t)$

5686 [1969, 835]. *Proposed by J. K. Washenberger, Virginia Polytechnic Institute*

The field $Q(t)$ of quotients of polynomials with rational coefficients is the standard example of a non-Archimedean ordered field. In this field $p(t)/q(t)>0$ if the leading coefficient of the product $p(t)q(t)$ is positive. Characterize the Archimedean ordered subfields of $Q(t)$.

Solution by G. A. Heuer, Concordia College. The only Archimedean ordered subfield of $Q(t)$ is Q. For, let F be such a subfield. Then Q, the prime field of

$Q(t)$, is a subfield of F. If $p(t)$ has larger degree than $q(t)$, then $p(t)/q(t)$ is larger than every element of Q, so neither $p(t)/q(t)$ nor $q(t)/p(t)$ can be in F. If $p(t)$ and $q(t)$ have the same nonzero degree, then $p(t)/q(t)=c+r(t)/q(t)$, where c is in Q and $r(t)$ has smaller degree than $q(t)$. Since $r(t)/q(t)$ is not in F, neither is $p(t)/q(t)$. Thus $F=Q$.

Also solved by J. W. Andrushkiw, John Bryant & Robert Gilmer, R. R. Guenther, Melvin Henriksen, M. L. Laplaza (Puerto Rico), O. P. Lossers (Netherlands), M. D. Mavinkurve (India), and the proposer.

REVIEWS

EDITED BY KENNETH O. MAY

COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER

Printed materials for review should be sent to Kenneth O. May, American Mathematical Monthly, Department of Mathematics, University of Toronto, Toronto 5, Canada. Films and correspondence relating to films should be sent to Seymour Schuster, Carleton College, Northfield, Minnesota 55057.

All unsigned material is written by the editors. A boldface capital **C** *in the margin indicates that a review is based in part on classroom use. Professors willing to write such a review should first inform the editor in order to avoid duplication.*

Introduction to Commutative Algebra. By M. F. Atiyah and I. G. Macdonald. Addison-Wesley, Reading, Mass., 1969. 137 pp. \$7.50. (Telegraphic Review, October 1969.)

This is a short 126 page introduction to commutative ring theory. It assumes the student has had a very solid first course (graduate level in North America, Part II level in England) in general algebra. Few specific results are assumed, but the maturity and sophistication that comes from such a course are assumed. The material is approximately what one might cover in a short course on ideal theory in commutative rings. However, here the emphasis is more on modules and localization than on classical ideal theory. In no way is it a substitute for such texts as Zariski-Samuel *Commutative Algebra I, II*, or Bourbaki *Algèbre Commutative.*

The authors acknowledge that the subject has its origins in algebraic number theory and in algebraic geometry; but here the emphasis is algebro-geometric and a serious student of number theory will still have need to study such texts as Hecke: *Algebraischen Zahlentheorie.* The authors' main motivation is the preparation of the student for a systematic study of homological algebra, the first step for serious study of modern algebraic geometry and class field theory.

THE AMERICAN MATHEMATICAL MONTHLY

(FOUNDED IN 1894 BY BENJAMIN F. FINKEL)

THE OFFICIAL JOURNAL OF

THE MATHEMATICAL ASSOCIATION OF AMERICA, INC.

VOLUME 77 NUMBER 8

CONTENTS

(continued on inside cover)

OCTOBER 1970

NOTICE TO AUTHORS

Articles should be typewritten on good quality paper, triple spaced with wide margins. Submit the original and a duplicate; keep a complete copy for protection against loss. Please follow the format used in current issues of the MONTHLY. The manuscript must make the author's intent clear to the printer. A paper accepted for publication by the editor but not acceptable to the printer will be returned to the author for retyping. Please use the American Mathematical Society *Manual for Authors of Mathematical Papers* as your guide and please conform to correct English usage. A *Manual for Monthly Authors* is available on request.

Prospective MONTHLY authors are advised to consult the Statement of Policy in the January 1969 issue, p. 2. Specialized research is usually unsuitable. Backlog: Main Articles 9 months, Math. Notes 9 months, Research Problems 7 months, Classroom Notes 7 months, Math. Education 8 months.

EDITORIAL CORRESPONDENCE AND MAIN ARTICLES: to HARLEY FLANDERS, American Mathematical Monthly, Tel Aviv University, Ramat Aviv, Israel (see Notice, p. 555); NOTES, etc.: to the corresponding Associate Editor; ADVERTISING CORRESPONDENCE: to RAOUL HAILPERN, Mathematical Association of America, SUNY at Buffalo, Buffalo, N. Y. 14214; CHANGE OF ADDRESS and SUBSCRIPTIONS: to A. B. WILLCOX, Mathematical Association of America, 1225 Connecticut Ave., N.W., Washington, D. C. 20036.

Annual dues for members of the Association (including a subscription to the American Mathematical Monthly) are $10.00. For nonmembers the subscription price is $15.00.

PUBLISHED BY THE ASSOCIATION at Washington, D. C., and Menasha, Wisconsin during the months of January, February, March, April, May, June–July, August–September, October, November, December.

Second-class postage paid at Washington, D. C., and additional mailing offices. Acceptance for mailing at special rate of postage provided for in the Act of February 28, 1925, embodied in Paragraph 4, Section 538, P. L. and R., authorized April 1, 1926.

PRINTED IN THE UNITED STATES OF AMERICA

SUMS OF SQUARES

OLGA TAUSSKY, California Institute of Technology

1. Introduction. Sums of squares is a major concept in mathematics going back to ancient days and yet of great current interest. It is a subject which links many different branches of mathematics and produces results which have a certain similarity but whose complete connection is still not understood. In recent years logicians have been much interested in the subject too. There are applications to and from logic to sums of squares. Statistics from its beginnings has been involved in sums of squares.

This article describes some of these ideas, but is by no means comprehensive. In particular the chapter on number theory is very incomplete. The presentation is a spotlight treatment, sometimes putting very deep results next to easier ones, although the latter may have a particular appeal and even importance. Proofs are included only when they are very brief. Some new ideas are incorporated.

This account is devoted to algebra and number theory on the whole, apart from describing facts which link up with analysis and topology and ought not to be separated. But sums of squares also occur in the very definition of the Hilbert space and all its consequences, e.g. in Parseval's theorem, in the definition of L^2-convergence, in normed algebras and such like. The composition of infinite quadratic forms will not be discussed. Not even the theory of finite euclidean and unitary space will be included, nor facts concerning orthonormal vectors, nor the theory of norms of finite matrices. Hence orthogonal and unitary matrices are not treated, nor automorphs of quadratic forms.

2. Pythagorean triangles and Fermat's last theorem. The first sum of squares we meet in our life is in Pythagoras' theorem

$$a^2 + b^2 = c^2, \tag{1}$$

where a, b are the sides of a right angled triangle with hypotenuse c. Later we meet in elementary trigonometry

$$\cos^2 \alpha + \sin^2 \alpha = 1 \tag{2}$$

and much later we meet

$$\cos^2 z + \sin^2 z \equiv 1 \tag{3}$$

The author received her training under P. Furtwängler at the University of Vienna, held assistantships at Göttingen and Vienna, and a fellowship at Girton College, Cambridge, and was a lecturer at London University. She was Scientific Officer with the British Aircraft Production Ministry and served with the Department of Scientific and Industrial Research, London. After many years with the National Bureau of Standards, she transferred to Cal Tech where she is Research Associate (equivalent rank of Full Professor). Leaves of absence were spent at NYU, Bryn Mawr College, Vienna (as Fulbright Professor), and UCLA.

Olga Taussky's prolific research centers on algebraic number theory, quaternions, and matrix theory. She was co-editor of Hilbert's *Gesammelte Abhandlungen* and edited several volumes of the Nat. Bur. Stand. Applied Math. Series. *Editor.*

for all complex values z. The Pythagorean triangles too turn up at an early stage in our education. They are right angled triangles whose sides have integral ratios like

$$3, 4, 5;\ 5, 12, 13; \cdots.$$

They are already mentioned in an old-Babylonian text discovered by Neugebauer and Sachs. It is known that there are infinitely many such triangles and that they are obtained from a parametric formula

$$\lambda(m^2 - n^2), \quad \lambda 2mn, \quad \lambda(m^2 + n^2) \tag{4}$$

with λ, m, n any integers. Sometimes the same triangle can be obtained several times by this formula, e.g. the triangle 6, 8, 10 is given by $\lambda=1$, $m=3$, $n=1$, and by $\lambda=2$, $m=2$, $n=1$. That every expression (4) leads to a Pythagorean triangle is clear. The converse will now be proved. We start with some general remarks.

At least one of a, b is even. For, if $a=2n+1$, $b=2m+1$, n, m integers, then $c^2=4N+2$, N an integer. This is impossible, for the square of an odd number is always of the form $4r+1$, r an integer, and that of an even number is divisible by 4. We shall assume that b is even. If a and b have a common factor $d\neq 2$, then $d|c$. Hence a/d, b/d also define a Pythagorean triangle. If, however, $d=2$ and $b/2$ is odd then we cannot remove it, unless $a/2$ is even, in which case we interchange the role of a and b. If, however, both $a/2$ and $b/2$ are odd then $(c/2)^2$ would again be of the form $(4M+2)$, M an integer, which is impossible.

The following elementary proof for the converse uses the fact that the expressions (4) suggest the formulas for $\cos 2\alpha$, $\sin 2\alpha$. Let

$$a^2 + b^2 = c^2.$$

Define the angle $\alpha (0 < 2\alpha < \pi/2)$ by

$$\frac{a}{c} = \cos 2\alpha, \quad \frac{b}{c} = \sin 2\alpha.$$

Since $\cos 2\alpha = \cos^2\alpha - \sin^2\alpha$ and $\cos^2\alpha + \sin^2\alpha = 1$ we have

$$\cos^2 \alpha = \frac{1}{2}\left(1 + \frac{a}{c}\right) = r_1, \text{ a rational,}$$

$$\sin^2 \alpha = \frac{1}{2}\left(1 - \frac{a}{c}\right) = r_2, \text{ a rational.}$$

Since $\sin 2\alpha = 2 \sin \alpha \cos \alpha$ we have

$$\sin \alpha \cos \alpha = \frac{1}{2}\,\frac{b}{c} = r, \text{ a rational.}$$

Hence

$$r = \sqrt{r_1 r_2}.$$

Hence

$$\sin \alpha = \frac{r}{r_1} \cos \alpha = \frac{n}{m} \cos \alpha,$$

where n, m are integers which we may suppose without a common factor. Put

$$\frac{\cos^2 \alpha}{m^2} = \lambda_0 \quad \text{(a rational) so that } \sin^2 \alpha = \lambda_0 n^2.$$

This gives

$$a = \lambda(m^2 - n^2), \qquad b = \lambda \cdot 2mn,$$

where $\lambda = \lambda_0 c$. We claim that λ is necessarily integral. For, if λ were fractional with p, a prime, dividing the denominator, then $p \mid (m^2 - n^2)$, $p \mid 2mn$. If $p \neq 2$, then $p \mid m$ or $p \mid n$. Since $p \mid (m+n)$ or $p \mid (m-n)$ it follows that $p \mid m$ and $p \mid n$ which is a contradiction. If $p = 2$ and $p \nmid m$, $p \nmid n$ then b is not even as was assumed.

In contrast to the various elementary proofs of (4) a proof using Galois theory will now be given. It is based on Hilbert's Theorem 90 which concerns algebraic extension fields with a cyclic Galois group. This theorem is obtained nowadays as a special case of a theorem in Galois cohomology (see e.g. Jacobson, Algebra III.)

Let F be a cyclic extension of a field K of relative degree l. Let S be a generator of the Galois group of F over K. For any $\alpha \in F$ we write α^S for the automorphism defined by S. We then have

$$\operatorname{norm}_{F/K}(\alpha) = \operatorname{norm}_{F/K}(\alpha^S).$$

Hence by the multiplicativity of the norm we have

$$\operatorname{norm}_{F/K}(\alpha/\alpha^S) = 1.$$

Hilbert's theorem states that, conversely, any element $\beta \in F$ with norm $\beta = 1$, is of the form $\beta = \alpha/\alpha^S$ for a suitable $\alpha \in F$.

Apply the theorem to the situation where F is the extension obtained from the rational number field Q by adjoining $\sqrt{-1}$, i.e., the set of elements $m + in$, $m, n \in Q$. This field is cyclic with respect to Q and of degree 2. Further

$$(m + in)^S = m - in \quad \text{and} \quad \operatorname{norm}(m + in) = m^2 + n^2.$$

Let then $a^2 + b^2 = c^2$ hold with a, b and c in Z, the ring of integers. This implies

$$\operatorname{norm}\left(\frac{a}{c} + i\,\frac{b}{c}\right) = 1.$$

By Hilbert's theorem

$$\frac{a}{c} + i\,\frac{b}{c} = \frac{m + in}{m - in} = \frac{(m + in)^2}{m^2 + n^2}$$

for some $m, n \in Z$. Comparing the real and imaginary parts (4) emerges.

We add three further comments to the study of Pythagorean triangles:

(1) The expressions (4) can be given another interpretation: It is clear that Pythagorean triangles have much in common with complex numbers. The product of two complex numbers m_1+in_1, m_2+in_2 is $m_1m_2-n_1n_2+i(m_1n_2+m_2n_1)$. This associates two bilinear forms $m_1m_2-n_1n_2$, $m_1n_2+m_2n_1$ with the field of complex numbers. The corresponding quadratic forms

$$m^2 - n^2, \; 2mn$$

are exactly the expressions (4).

Similarly, one can associate a set of n bilinear (respectively quadratic) forms with any basis of an algebra. This idea is being investigated separately, particularly for algebraic number fields.

(2) The Pythagorean triangles form a group under a certain composition law: more precisely to every triangle a, b, c consider the whole set λa, λb, λc, with $\lambda = 1, 2, \cdots$, as an element of the group under consideration. Further identify all the four pairs $\pm a$, $\pm b$. We may therefore assume a, b as coprime positive integers. Exactly one of these two integers is then even, because $a^2+b^2=c^2$ is a square, hence cannot be $\equiv 2(4)$. We will assume that b is an even number.

Let a_i, b_i, $i=1, 2$, be a pair which generate a Pythagorean triangle. Then it follows at once that

$$A = a_1a_2 + b_1b_2, \qquad B = a_1b_2 - a_2b_1$$

again generate a Pythagorean triangle. The set of triangles generated by A, B is the "product" of a_1, b_1 and a_2, b_2. If $a_1=a_2$, $b_1=b_2$, we obtain

$$A = a_1^2 + b_1^2, \qquad B = 0$$

which is equivalent with $A=1$, $B=0$. We consider this as the unit element in our group. For, this element when composed with a, b gives a, $-b$ which is equivalent with a, b. The associative law too follows for our composition if we again allow the above identification.

(3) The two quadratic forms $f=x^2-y^2$, $g=2xy$ associated with Pythagorean triangles, have a special property. Let

$$f_i = x_i^2 - y_i^2, \quad g_i = 2x_iy_i, \qquad i = 1, 2.$$

Then the pair of forms

$$f_1f_2 - g_1g_2, \qquad f_1g_2 + f_2g_1$$

are again the same forms f, g, but applied to the indeterminates:

$$x_1x_2 - y_1y_2, \qquad x_1y_2 + x_2y_1.$$

For $x_1=x_2$, $y_1=-y_2$ this gives the well-known relation $(x^2-y^2)^2+4x^2y^2 = (x^2+y^2)^2$.

The existence of these triangles made it desirable to know whether the equation

$$x^n + y^n = z^n \tag{5}$$

can be solved in integers x, y, z for $n>2$, apart from trivial cases. The not yet established statement that this is impossible is referred to as Fermat's last theorem. A similar question was raised concerning the relation (3) namely, do there exist two entire functions $f(z)$, $g(z)$, neither of them a constant, such that for some $n>2$

$$(f(z))^n + (g(z))^n \equiv 1. \tag{6}$$

A very brief proof was given by Iyer showing that such a pair does not exist. The equation (6) above is identical with

$$\prod(f(z) + \zeta g(z)) \equiv 1$$

when the product is taken over all solutions ζ of the equation $x^n+1=0$. Since none of the n factors can vanish the meromorphic function $f(z)/g(z)$ cannot assume any of the n values of ζ which would be in contradiction with the 'big' Picard theorem.

3. Sums of squares in number theory. This is an enormous subject which can only be touched briefly here. Again a very old theorem comes to our mind immediately: every positive integer is a sum of four squares of integers. Then we have the characterizations of integers which are sums of three squares and the well-known fact that an integer is a sum of two squares if and only if its square-free part is the product of primes $\equiv 1(4)$.

The quadratic form $x_1^2+x_2^2+x_3^2+x_4^2$ is not the only form $\sum a_i x_i^2$, a_i positive integers, which represents all positive integers. Forms of this nature are called universal and have been studied, e.g. by L. E. Dickson, Kloosterman, Linnik, Pall, and Ramanujan. Forms which represent all but one positive integer were examined by Halmos. Heilbronn showed that there exist four continuous functions $f_i(x)$ such that every rational x is represented by $\sum f_i(x)^2$. To represent positive integers by more than four, in particular by many, squares has been studied too and the function giving the number of representations of the fixed integer n as the sum of exactly s squares has been of much interest for a long time. Recently Bateman pointed out that the function $f_s(n)=(2s)^{-1}r_s(n)$ is multiplicative precisely for $s=1, 2, 4, 8$, where $r_s(n)$ is the number of representations of n as a sum of s integral squares.

The problem of representing algebraic integers as sums of squares in the same field has been much studied, but will not be discussed here.

Next we look at *integral* symmetric matrices, a subject not fully investigated so far.

(The fact that a *real* symmetric matrix has real characteristic roots links their study with sums of squares, but this aspect will be discussed in Chapter 6.)

Even the subject of rational symmetric matrices is not fully explored. For 2×2 matrices the characteristic roots of such a matrix must lie in a field $Q(\sqrt{m})$, m a sum of two squares. This can be checked easily from the formula for the zeros of the (quadratic) characteristic polynomial. However, it can also be obtained as a special case of the following fact: Let F be an extension of degree n of the rationals and let it have a symmetric $n\times n$ Q-representation. Then this is equivalent with the existence of n elements $\alpha_1, \cdots, \alpha_n$ in F such that $\sum_i \alpha_i^{(r)}\alpha_i^{(s)}=0$, $r, s=1, \cdots, n$, $r\neq s$, where the upper suffices denote the conjugate elements with respect to Q. First we show that this is necessary: Let A be the Q-matrix which represents the primitive element α. Since A is symmetric its characteristic vectors are orthogonal. Since A is a Q-matrix these vectors can be chosen in $Q(\alpha^{(i)})$, $i=1, \cdots, n$, and as the conjugates of the vector corresponding to the characteristic root α. Hence the components of this vector are of the form of the above $\alpha_1, \cdots, \alpha_n$. Next we show sufficiency: the matrix

$$(\alpha_i^{(j)})\begin{pmatrix} \alpha & & & \\ & \alpha^2 & & \\ & & \ddots & \\ & & & \alpha^{(n)} \end{pmatrix}(\alpha_i^{(j)})'$$

is rational and symmetric; here and later the prime indicates the transpose. Hence we have obtained a rational symmetric representation of $Q(\alpha)$.

From the orthogonality it also follows that F is totally real.

The case $n=2$ gives an alternative proof for a fact mentioned above, for the two elements α_1, α_2 are expressible as

$$\alpha_1 = a + b\sqrt{m}, \qquad \alpha_2 = c + d\sqrt{m},$$

where a, b, c, d, m are in Q and m is not a square in Q. Then

$$a^2 - mb^2 + c^2 - md^2 = 0.$$

Hence

$$a^2 + c^2 = m(b^2 + d^2).$$

Symmetric $n\times n$ matrices over the integers with a given characteristic polynomial of degree n, with integer coefficients and 1 as coefficient of x^n can sometimes best be studied by looking first for general matrices, i.e. not necessarily symmetric ones. Such matrices fall into classes: two matrices being considered equivalent if they belong to the same integral unimodular similarity class. For irreducible polynomials this leads to an integral $n\times n$ representation of the ring generated by a zero of the polynomial and hence to a rational $n\times n$ representation for the algebraic number field generated by it. Let A be a suitable matrix. Then A', the transpose, is a matrix-zero of the same polynomial. Under special circumstances it can belong to the class of A. We then have

$$A' = S^{-1}AS$$

when S is integral and unimodular. If in addition S is p.d. (positive definite) and even of the form $S=TT'$ with T integral (this follows from p.d. for $n\leqq 7$, (see Chapter 4)), then

$$T^{-1}AT = (T^{-1}AT)'$$

Hence the class of T contains a symmetric matrix and conversely. Faddeev, Shapiro, Bender studied symmetric matrices over algebraic number fields with given characteristic polynomials. However, not all polynomials which can turn up for symmetric matrices have been characterized so far.

This whole chapter belongs partly to the theory of positive definite matrices. They are treated in the next chapter which also includes further number theoretic results. Also the similarity between a matrix and its transpose and the connection with symmetric matrices will turn up there again for the case of the real number field.

4. Positive definite (p.d.) matrices. These are real symmetric matrices with positive characteristic roots. Again a very old fact is our starting point: a positive definite quadratic form with real coefficients is a sum of squares of linear forms. Positive definite hermitian forms are expressible as $\sum l_i(x)\bar{l}_i(\bar{x})$, where $l_i(x)$ is a linear form and $\bar{l}_i(x)$ is the form with the conjugate coefficients.

For the matrix itself this means in the real case that it can be factorized as AA' where A' is the transpose of A and in the complex case as BB^* where $B^*=\bar{B}'$ is the transposed complex conjugate.

One of the most important uses of p.d. matrices is to generalize facts in many different branches of mathematics, by replacing the identity matrix by a given p.d. matrix. We begin with two specific examples.

(1) *The orthogonal matrices.* They leave $\sum x_i^2$ unchanged. This was generalized in the study of the 'automorphs' of p.d. quadratic forms.

(2) *The field of values of a complex $n\times n$ matrix A.* It is the set of numbers in the complex plane given by x^*Ax/x^*x, where $x\neq 0$ is an arbitrary complex n-vector. Givens introduced the generalized field of values x^*AHx/x^*Hx when H is a p.d. form. This was recently even extended to operators.

In differential geometry Riemannian geometry extends Euclidean geometry. In the subject of partial differential equations the theory of elliptic equations extends that of the Laplace equation. There are many examples in the calculus of variations. There are examples in number theory on all levels.

We begin with the discussion of products of real symmetric matrices. It can be shown that every real matrix is the product of two symmetric matrices—more generally a matrix with elements in an arbitrary field F can be expressed as the product of two symmetric matrices in the same field. In the case of the reals: if one of the two symmetric factors is positive definite then the product is similar to a real diagonal. If both factors are p.d. then the product is similar to a p.d. diagonal. Connected with this is the following fact: while every matrix with elements in a field F is similar to its transpose via a symmetric matrix

over F, the latter can be chosen in the form AA' (and therefore p.d. for the reals) if the original matrix is similar to a symmetric matrix, (i.e. has real characteristic roots and is diagonalizable in the case of the reals). Real matrices with positive determinant can always be expressed as products of p.d. matrices, in fact only four factors are needed, unless the matrix is a negative scalar in which case five factors are required, in general. This was shown by Ballantine who also characterized products of three p.d. factors. Products of two p.d. matrices had been characterized by Taussky.

Positive definite matrices can be employed to determine the signs of the real parts of the characteristic roots of a general matrix. This is an important practical problem. By a theorem of D. C. Lewis, Jr., one can find for any matrix A with simple elementary divisors a p.d. hermitian G such that the roots λ of $\det(GA+A^*G-2\lambda G)=0$ are the real parts of the characteristic roots of A. Also there is the matrix version of Lyapunov's stability criterion:

A matrix A is stable if and only if a p.d. G exists such that $AG+GA^*$ is negative definite.

This was generalized to give statements concerning the signs of the real parts of the characteristic roots.

The fact that every real symmetric matrix can be transformed to diagonal form by an orthogonal matrix can be generalized by saying that a pair of symmetric matrices, one of which is p.d., can be transformed to diagonal form simultaneously. This again is generalized by saying that a pair of symmetric matrices S_1, S_2 which generate a pencil $\lambda S_1+\mu S_2$ which contains a p.d. matrix can be simultaneously diagonalized. Such pencils have been studied recently and were linked up with the convex cone formed by the p.d. $n\times n$ matrices in the $n(n+1)/2$ dimensional space.

The cone of p.d. matrices H is invariant under the transformation

$$AHA'$$

when A is any non-singular matrix of the same dimension. Thus they form a 'positivity domain' like the positive vectors which are invariant under transformation by a positive matrix (in this case even a non-negative ($\neq 0$) matrix). The transformation defined by A on the linear space of symmetric matrices can be regarded as a 'positive operator' and the finite version of the Krein-Rutman theorem concerning such operators can be applied to it. From this it follows that the matrix which corresponds to the operator has a positive dominant characteristic root. The matrices H which are 'characteristic vectors' are of interest too.

We now turn to consideration of number theory and we mention some facts concerning the factorization XX' and the problem of representing integers by p.d. forms.

Positive definite matrices play a big role in number theory. In particular, unimodular integral matrices A have been studied by Hermite and Minkowski who showed that for $n\leqq 7$ every such matrix is of the form BB', with B an

integral $n \times n$ matrix. For $n=8$ this is not any longer true as an example by Korkine and Zolotareff shows. Mordell showed that there are two classes in this case if we count two matrices A, B as belonging to the same class if $B=SAS'$ where S is an integral and unimodular matrix. The number of classes is finite in all cases.

A special case of p.d. unimodular matrices arises from the set of group matrices for the finite group G (a ring isomorphic with the integral group ring of G). These matrices are of the form $(a_{rs^{-1}})$, where r, s range over the elements of G in a fixed order. The unimodular ones correspond to the units in the group ring. A factorization of the above type is possible only if B is a group matrix for the same group, times an integral matrix P with $PP'=I$. For $n \leqq 7$ such a factorization is possible always, for $n=8$ there are again two classes. For $n=9$, however, there is only one class. These results were obtained by Taussky, M. Newman, R. C. Thompson, M. Kneser, some still unpublished.

The results concerning the factorization of positive definite unimodular matrices have been extended to hermitian matrices and to matrices over quadratic number fields. For instance, any matrix

$$A = \begin{pmatrix} a & \alpha \\ \bar{\alpha} & b \end{pmatrix}$$

with a, b positive integers, $\alpha = x+iy$ a Gaussian integer and $ab-\alpha\bar{\alpha}=1$ can be factorized into BB^* with B a square matrix of Gaussian integers. This result leads to an alternative proof for Lagrange's theorem on expressing a positive integer a as a sum of four squares. The following fact is needed: Given a number a, integers $b>0$, x, y can be found such that

$$x^2 + y^2 = -1 + ab.$$

Starting with the positive integer a the matrix A is now determined. The factorization $A=BB^*$ then expresses a as a sum of four squares. (This was pointed out by M. Newman during a discussion with the author.)

Even if the p.d. integral matrix is not of the form XX', X an integral square matrix, it may still be expressible in this form with X a rectangular matrix with more columns than rows. For the corresponding quadratic form this means that it is still expressible as a sum of squares of integral linear forms, but with more than n terms. However, even this is not always possible. This was studied by Mordell, Erdös, Chao Ko, Pall and Taussky.

Alternatively, the problem of expressing a matrix A in the form XBX' has been studied in full generality by Siegel, where now A and B do not need to have the same dimension. Hence this includes the representation of numbers by p.d. quadratic forms.

The extension of the problems to algebraic number fields led Siegel to a conjecture which was established quite recently. It concerns 'classes' and 'genera' of forms. It is known that over the rationals the forms

$$\sum_1^4 x_i^2, \qquad \sum_1^8 x_i^2$$

are in a genus of one class. Over totally real fields the form $\sum_1^4 x_i^2$ lies in a genus of one class only for $Q(\sqrt{2})$, $Q(\sqrt{5})$. This is what Siegel had conjectured. After some initial progress by Dzewas it was established by Barner.

5. Formally real fields. The congruence

$$x^2 + y^2 \equiv -1(p), \qquad p \text{ any prime},$$

used in Chapter 4, can be used as a link with this theory since the integers modulo a prime p form a field. A formally real field is characterized by the fact that -1 cannot be expressed as a sum of squares in that field. The fact that the above congruence can be solved for any p not only shows that the residues mod p do not form a formally real field (a fact obvious from the finite characteristic), but it gives the actual expression for -1.

If the field is not formally real and if its characteristic $\neq 2$ then every element in the field is a sum of squares.

If -1 is a sum of squares then it is of interest to study the minimum number of terms in the representation of -1 for various fields F. It is easy to see that 3 cannot occur as a minimum. For, assume that 3 is the minimum and let

$$-1 = x_1^2 + x_2^2 + x_3^2, \qquad x_i \in F, \quad x_i \neq 0.$$

This implies

$$0 = x_0^2 + x_1^2 + x_2^2 + x_3^2, \qquad x_i \in F.$$

Hence

$$\begin{aligned} 0 &= (x_0^2 + x_1^2)^2 + (x_0^2 + x_1^2)(x_2^2 + x_3^2) \\ &= (x_0^2 + x_1^2)^2 + (x_0x_2 - x_1x_3)^2 + (x_0x_3 + x_1x_2)^2, \end{aligned}$$

so that transferring $(x_0^2+x_1^2)^2$ to one side and dividing across by it (using the fact that $x_0^2+x_1^2=0$ by assumption) we get a representation of -1 as a sum of two squares. This contradicts our assertion. This proof depends on the multiplication of complex numbers and their norms. A similar proof using the multiplication of quaternions and the multiplicativity of their norms shows that 5, 6, 7 cannot occur as a minimum. The same idea, using Cayley numbers, shows that the numbers $9, \cdots, 15$ cannot occur as a minimum. The question concerning the possible minima had been raised by van der Waerden in 1932 and was settled only quite recently by Pfister. He showed that only powers of 2 can occur as a minimum and that every such power does occur for some field. Pfister uses results of Cassels on quadratic forms for his proof. The relevant theorems are:

Let F be a field of characteristic $\neq 2$. Let $d \in F$ and x be an indeterminate. Necessary and sufficient for x^2+d to be a sum of $n>1$ squares in $F(x)$ is that either -1 or d is a sum of $n-1$ squares in F.

Let R be the field of real numbers and let $x_1, \cdots, x_n$ be indeterminates over R. Then $x_1^2+ \cdots +x_n^2$ is not a sum of $n-1$ squares in $R(x_1, \cdots, x_n)$.

As the cases 1, 2, 4, 8 show, the result is connected with the composition of sums of squares which will be discussed in the next chapter.

By defining 'positive' in real fields as 'sums of squares' an ordering can be introduced in such fields.

Symmetric matrices over formally real fields have been studied. The set of all their characteristic roots form a field which is real closed. Krakowski and recently also Bender studied symmetric matrices over arbitrary fields with given minimum polynomial.

An application of formally real fields appeared in a very unexpected connection: R. C. Thompson proved (generalizing a theorem by Shoda obtained for algebraically closed fields) that, with the exception of certain 2×2 matrices over GF(2), every unimodular matrix A with elements in a field F is a commutator $B^{-1}C^{-1}BC$ in F. He also studied the question: when can the factors B, C themselves be chosen unimodular? For the case that A is a scalar matrix, this depends, among other things, on whether -1 is a sum of two squares in F. Later Thompson examined the case when the factors B, C have given determinants b, c and then the representation of -1 in the form bx^2+cy^2 becomes critical.

6. Composition of sums of squares, anticommuting matrices, composition algebras. Hurwitz showed that $n=1, 2, 4, 8$ are the only values of n for which identities of the following type hold:

$$\sum_1^n x_j^2 \sum_1^n y_k^2 = \sum_1^n [l_i(x, y)]^2,$$

where l_i are bilinear forms in the x_j, y_k. Pfister's results concerning -1 as a sum of squares are linked with an extension of this problem: he allows the $l_i(x, y)$ to be *rational* functions. In this way he obtains an identity for any n which is a power of 2. For $n=8$ such an identity had been obtained independently by Taussky by a different method and this result was extended to $n=16$ by Eichhorn and Zassenhaus:

The usual way to obtain the above mentioned identities is from the product $\alpha\beta$ of two complex numbers, respectively quaternions, or Cayley numbers, and using the fact that norm $\alpha\beta=$norm α norm β in all these cases. The identities found by Taussky and Eichhorn and Zassenhaus were, however, derived for $n=2$ from the reals, for $n=4$ from the complex field, for $n=8$ from quaternions, for $n=16$ from Cayley numbers. The method is based on the generalization of the relation $\det X \,\overline{\det X} = \det XX^*$ when X^* means the transpose conjugate.

The identities are special cases of Gauss' concept of composition of quadratic forms: two n-ary quadratic forms $f(x_i)$, $g(x_i)$ are said to permit composition if the product fg can again be expressed as a quadratic form under a bilinear transformation of the indeterminates.

Hurwitz' proof is based on matrix theory and leads to the enumeration of skew symmetric $n\times n$ matrix pairs which are anticommuting. Such pairs are of interest in many connections and will be mentioned again in Chapter 7. Using an idea of Jordan, von Neumann, and Wigner the theory of group representations was employed by Eckmann (after he replaced Hurwitz' matrices by abstract elements and -1 by an element of order 2) to give a proof of Hurwitz' theorem.

Freudenthal uses a projective geometry over the field of two elements and Desargues' theorem to prove Hurwitz' theorem and Chevalley uses Clifford algebras. An account of this and connected ideas are in a paper by van der Blij.

Later Albert, Kaplansky, and Jacobson imbedded the problem into the study of composition algebras. We start by defining a normed algebra. Let $e_1, \cdots, e_n$ be a basis for the algebra. Let

$$a = a_1e_1 + \cdots + a_ne_n$$

be an element of the algebra. Define

$$\text{norm}\, a = |a| = \sum a_i^2.$$

The algebra will be called normed, if

$$|ab| = |a||b|.$$

If the algebra is over the reals, has an identity and is normed then it can be shown that it is either the reals, the complex field, the quaternions or the Cayley numbers. This gives a proof for the Hurwitz theorem. Jacobson treated a more general situation. He starts with a quadratic form $N(x)$ defined on a vector space V over a field of characteristic $\neq 2$. He assumes that $N(x)$ permits composition, i.e., there exists a bilinear composition xy in V such that

$$N(x)N(y) = N(xy), \qquad x, y \in V.$$

The product xy makes an algebra out of V which is then called a composition algebra. However, Albert had shown that forms of dimension 2^n permit composition even for fields of characteristic 2.

The Hilbert identities

$$\left(\sum_1^r x_i^2\right)^m = \sum \rho_k(a_{1k}x_1 + \cdots + a_{rk}x_k)^{2m}$$

with ρ_k positive rationals and a_{ik} integers were used in Hilbert's original solution of the Waring problem, i.e., the proof of the following assertion: every positive integer is a sum of n-th powers of integers and the number of these is deter-

mined solely by n. A simpler proof of the identities goes back to Hausdorff. They have the flavour of a composition identity and could possibly be generalized for products

$$\sum x_i^2 \sum y_i^2 \cdots.$$

The associative algebras among the above-mentioned four algebras over the reals are characterized by other properties, e.g., Pontryjagin obtains them from topological properties of fields in which addition and multiplication are continuous functions under some topology. The Gelfand-Mazur theorem states that they are the only normed algebras which are also fields. They will be discussed further in the next chapter.

7. Division algebras over the reals, n-dimensional spheres, n-dimensional Laplace differential equations. Frobenius proved that the reals, the complex numbers and the quaternions are the only division algebras over the reals if commutativity is not required, but associativity retained. In spite of arduous attempts there is still no algebraic proof available for the fact that $n=1, 2, 4, 8$ are the only numbers of base elements for which real division algebras exist, if associativity is not required any longer. The only proofs available so far rely heavily on deep algebraic topology. This approach was initiated by H. Hopf and continued by Stiefel and others with a final break-through by Bott, Kervaire, Milnor, Adams.

The norm in the case of the best known division algebras, the complex numbers, quaternions, Cayley numbers, is the sum of the squares of the coordinates.

The role of these hypercomplex systems in almost any part of mathematics is well known. Quaternions have applications in most abstract parts of mathematics, but also in concrete ones from where they stem. They were introduced by Hamilton and are of great use in theoretical physics, as well as in ring theory, group theory and number theory. If the coordinates are permitted to be complex then the system is no longer a division algebra, but it has other applications. Generalized quaternions are in much use too, they are again algebras with four base elements, but the role of -1 in the products is taken over by other scalars.

In abstract group theory the quaternions are linked with the quaternion group which is the hamiltonian group of lowest order. A hamiltonian group is a non-abelian group, all of whose subgroups are normal. The real group algebra of the quaternion group is homomorphic with the real quaternions.

The norms of these hypercomplex systems link them immediately with the n-dimensional spheres. In particular the fact that no division algebra over the reals with three base elements exists is connected with the fact that the 3-dimensional sphere, the set of triples x_1, x_2, x_3 with $x_1^2+x_2^2+x_3^2=1$, cannot be made into a topological group. The latter result follows easily from Brouwer's fixed point theorem. E. Cartan proved that the n-dimensional sphere is a group space

only for $n=1, 2, 4$. It was the study of the topological properties of the higher dimensional spheres which led to the success in the investigation of the real division algebras. That the associative case follows easily from Cartan's result was noticed by Taussky. Recently a new proof of Cartan's result was given by M. Curtis and Dugundji. An earlier proof came from Samelson.

The ring $R[x_1, \cdots, x_n]$ of polynomials in n variables with real coefficients 'modulo the unit sphere', i.e. the ring

$$R[x_1, \cdots, x_n]/(x_1^2 + \cdots + x_n^2 - 1)$$

has been studied in several connections:

Swan showed that a 'unimodular' vector $(x_1, \cdots, x_n)$ over this ring cannot always be completed to a unimodular matrix, unless $n=1, 2, 4$, or 8.

Estes and Butts showed that for $n=3$ composition of quadratic forms is not possible in this ring.

Another subject connected with the classical division algebras is the Laplace differential equation

$$\left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \cdots + \frac{\partial^2}{\partial x_n^2}\right) u = 0.$$

For $n=2$ the Laplace equations for two functions of two real variables are a consequence of the Cauchy-Riemann equations. The question was raised (and answered) by Taussky as to whether an analogous situation exists for other values of n. It turns out this can happen only for $n=2, 4, 8$ and in fact it is a rather simple consequence of the result concerning the division algebras over the reals. However, a purely algebraic proof was obtained by Stiefel subsequently. The 4-dimensional case leads to a set of generalized Cauchy-Riemann equations which were much studied by Fueter and his school for the purpose of generalizing complex function theory. In slightly changed form they appear in theoretical physics as the Dirac equations.

Eichhorn, who had previously contributed to the study of generalized Cauchy-Riemann equations, considered the following more general problem recently: Let X be a vector space over a field F of characteristic $\neq 2$. Let $x \in X$. Consider linear mappings $L(x)$ of X into $\mathrm{Hom}(X, X)$ such that another such mapping exists so that

$$M(x)L(x) = \mu(x)I$$

when I is the identity mapping and $\mu(x) \not\equiv 0$ is a mapping of X into F. The function $\mu(x)$ can be interpreted as a quadratic form and if this form is positive definite and of rank n and X an n-dimensional vector space we are back in the problem of the generalized Cauchy-Riemann equations. However, it is now shown that for any form, whether p.d. or not, as long as it has full rank n, the only possible values of n are 1, 2, 4, 8. But also the case of lower rank $r<n$ is considered. For $n=p2^g$, p odd, Eichhorn obtains $r \leqq 2g+2$. The results are

based on the following result: Let $n=p2^g$, p odd. Let F be an arbitrary field of characteristic $\neq 2$. Let A_i, $i=1, \cdots, r-1$, be a set of $n\times n$ matrices with elements in F with the properties: $A_i^2=\alpha_i I$, $\alpha_i\in F$, $\alpha_i\neq 0$ and $A_iA_k+A_kA_i=0$, $i\neq k$. Then $r\leqq 2g+2$.

Connected with this fact is a result by Adams, Lax and Phillips: Let $A_1, \cdots, A_k$ be a set of real $n\times n$ matrices such that $\sum\lambda_iA_i$ is non-singular for all real λ_i, except $\lambda_1=\cdots=\lambda_n=0$. Let $n=2^{b+4c}(2a+1)$, $0\leqq b\leqq 3$. Then $k\leqq 2^b+8c$.

Anticommuting matrices are much connected with the various problems studied in this and the preceding chapter.

Dieudonné studied the following generalization of results by Eddington and by M. H. A. Newman. Let F be a not necessarily commutative field of characteristic $\neq 2$, V an n-dimensional right vector space over F, let σ be an automorphism of F and γ an element of F such that $\gamma^\sigma=\gamma$ and $\xi^\sigma=\gamma^{-1}\xi\gamma$ for $\xi\in F$. Determine the maximal number of semi-linear transformations u_k of V, relative to the automorphism σ, satisfying the relations

$$u_k^2(x) = x\gamma \quad \text{for } x \in V,\ \forall k$$
$$u_hu_k = -u_ku_h \qquad h \neq k.$$

Quaternions and Cayley numbers have many applications in number theory. Lipschitz had studied the ring of quaternions with rational integral coordinates, but Hurwitz later noticed that they do not form a maximal order. By order we understand a subring containing 1 and a basis for the algebra. He constructed the following basis for the maximal order: $(1+i+j+k)/2$, i, j, k. He then studied the factorization of rational primes in this ring. Similar problems have been studied for Cayley numbers by various authors, e.g. Coxeter, Lamont, Linnik, Mahler, Pall, Pall and Taussky, and Rankin.

If $e_1, \cdots, e_n$ ($n=4$, respectively 8) is a basis for an order and $x_1, \cdots, x_n$ indeterminates then

$$(x_1e_1 + \cdots + x_ne_n)(x_1\bar{e}_1 + \cdots + x_n\bar{e}_n)$$

is the norm form of the order if $\bar{e}_i$ is the conjugate of e_i. The problems associated with these forms can then be studied via the associated orders. Here the work of Brandt was basic and recently Kaplansky, Estes and Pall have made contributions.

8. Positive definite polynomials. This chapter is closely linked with the earlier chapters on number theory and on formally real fields. A positive rational number can be expressed as a sum of squares and Hilbert asked whether a real positive definite polynomial, i.e., one which assumes positive values only, can be expressed as a sum of squares of polynomials or, failing this, whether it can be expressed as a sum of squares of rational functions. He gave an example of a positive definite polynomial which cannot be expressed as a sum of squares of polynomials. Recently Motzkin gave a rather simple example

of such a polynomial, namely $(x_1^2+x_2^2-3x_3^2)x_1^2x_2^2+x_3^6$. He found this in connection with a study of inequalities in which he expresses the difference of the two sides as sums of squares. Other examples were found recently by R. Robinson:

$$x^2(x^2-1)^2+y^2(y^2-1)^2-(x^2-1)(y^2-1)(x^2+y^2-1)$$

$$x^2(x-1)^2+y^2(y-1)^2+z^2(z^2-1)^2+2xyz(x+y+z-2).$$

Artin solved Hilbert's question completely, showing that an expression by rational functions does indeed exist. Recently the subject was reactivated by asking for a quantitative result, namely the minimum number of terms and explicit representations. It was shown by Pfister that a definite rational function of n variables in a real closed field is a sum of 2^n squares in this field. This is only an upper bound, but for $n=2$ a smaller number will not suffice. Quite recently Cassels, Ellison and Pfister showed that the Motzkin polynomial is not a sum of 3 squares. The lower bound $n+1$ follows from Cassels' Theorem showing that $1+x_1^2+\cdots+x_n^2$ is not a sum of n squares. Ax had shown earlier that Artin's own work can be used to imply the bound 2^n if a further condition is fulfilled which he showed was in fact true for $n=3$.

For polynomials with integral coefficients the following facts have been studied: If the polynomial assumes square values for all integral arguments then $f(x)$ is itself the square of an integral polynomial. A much deeper result was obtained in a diophantine formulation by Siegel: An integral polynomial, not a square, can attain a square value for a finite number of integral values only. Recently LeVeque generalized the question: if $f(x)$ assumes only values which are sums of two squares, is $f(x)$ a sum of two squared polynomials? This was answered affirmatively by several authors.

9. Sums of squares in Galois theory. The converse problem of Galois theory, namely to find a normal algebraic extension with a given Galois group, leads in some special cases connected with the number 2 to sums of squares. The two following cases will be discussed:

(i) The cyclic group of order 4.

(ii) The quaternion group.

Let k be a given field and F a separable normal extension. In case (i) there exists exactly one quadratic field F_0 between k and F. This field is generated by the square root of an element $\mu\in k$. It can be shown that μ is a sum of two squares in k. Conversely, any sum of two squares occurs in this connection. There are various proofs for this. If μ is a sum of squares in k, say $\mu=\mu_1^2+\mu_2^2$, $\mu_i\in k$, then $F_0=k(\sqrt{\mu_1^2+\mu_2^2})$ has the property that $-1=\text{norm}_{F_0/k}(\rho)$ where $\rho\in F_0$. For, every element in F_0 is of the form $\alpha+\beta\sqrt{\mu_1^2+\mu_2^2}$, α, $\beta\in k$ and the norm of this element is $\alpha^2-\beta^2(\mu_1^2+\mu_2^2)$. This can be made equal to -1 for $\alpha^2=\mu_1^2/\mu_2^2$ and $\beta^2=1/\mu_2^2$. Conversely, if $-1=\text{norm}_{F_0/k}(\rho)$, $\rho\in F_0$, then F_0 is generated by the square root of a sum of two squares. For $k=Q$ it can happen that -1 is even the norm of a unit.

In case (ii) the field F contains three quadratic fields F_1, F_2, F_3 between k

and F. Let F_i be generated by $\sqrt{\mu_i}$, $\mu_i \in k$. Then $\mu_1\mu_2 = \mu^2\mu_3$ where $\mu \in k$. It can be shown by elementary field theory that each μ_i can be represented as a sum of three squares in k. However, two expressions which are sums of three squares do not in general have a product with the same property. Hence μ_1, μ_2 cannot be arbitrary sums of three squares. Pairs of sums of three integers whose product is of the same type can be characterized easily from the known characterization of such integers. Not all such pairs, however, qualify for quaternion fields. A parametric representation for μ_1, μ_2 was given by G. Bucht. The 'sum of three squares' character of such a representation will now be explained (The following treatment is due to Cassels arising out of a discussion with the author.):

The elements μ_1, μ_2 can be obtained as:

$$\mu_1 = \frac{u^2l^2 + s^2u^2 + s^2m^2}{l^2r^2 + l^2v^2 + s^2v^2} = \frac{u^2(l^2 + s^2) + s^2m^2}{v^2(l^2 + s^2) + l^2r^2}$$

$$\mu_2 = \frac{u^2r^2 + m^2r^2 + m^2v^2}{l^2r^2 + l^2v^2 + s^2v^2} = \frac{m^2(r^2 + v^2) + u^2r^2}{l^2(r^2 + v^2) + s^2v^2}.$$

It can be shown, conversely, that any pair of elements $\mu_1, \mu_2 \in K$ for which these relations hold have the property that $k(\sqrt{\mu_1}, \sqrt{\mu_2})$ can be extended to a quaternion field. That μ_1 is a sum of three squares follows from the fact that both denominator and numerator are norms in the field generated by $i\sqrt{l^2+s^2}$, hence their quotient is again such a norm, hence clearly a sum of 3-squares. An analogous fact is true for μ_2. Next we show that $\mu_1\mu_2$ is a sum of three squares. Since μ_1, μ_2 have the same denominator we need only worry about the numerators. Both of them are norms of the field generated by $i\sqrt{u^2+m^2}$; hence their product is a norm in this field. Finally, we give an example showing that not every set μ_1, μ_2, $\mu_1\mu_2$, all sums of three squares, comes in question. Take $\mu_1 = 3\cdot 73$, $\mu_2 = 3\cdot 37$. Then

$$\mu_1 v^2 + \mu_2 l^2 = u^2 + m^2,$$

hence

$$(*) \qquad 3(73v^2 + 37l^2) = u^2 + m^2.$$

This implies that $u \equiv m = 0(3)$, hence $v^2 + l^2 \equiv 0(3)$, hence $v \equiv l \equiv 0(3)$. Remove a factor 3 from u, m, v, l in (*) and repeat the process. This will finally lead to a contradiction.

A proof for the fact that the μ_i are sums of three squares by non-elementary methods was given by Reichardt.

10. Rational arctangents.*

We now return to a problem concerning the sums of two squares, which, in principle, goes back to Gauss.

* This chapter was written by John Todd.

In elementary trigonometry one encounters relations of the following form:

$$\arctan 239 = 4 \arctan 5 - 5\pi/4$$
$$\arctan 99 = \arctan 12 - 2 \arctan 5 - \arctan 2 + 5\pi/4.$$

We take up the question of generating all such relations or rather a basis for them. (One of the reasons for the study of those relations is to find convenient methods of calculating π.)

We shall write (x) for that value of arctan x between 0 and $\frac{1}{2}\pi$, so that, in particular, $(1)=\pi/4$. We ask, to begin with, can we have relations of the form

$$(2) = r(1)$$
$$(3) = s(1) + t(2),$$

where r, s, t are integers?

Suppose the first relation holds. Then the complex numbers $1+2i$ and $(1+i)^r$ necessarily have the same argument so that their ratio

$$(1 + 2i)/(1 + i)^r$$

is necessarily real. Since the real part of the denominator is an integer, m say, it follows that this ratio must be $1/m$. If we take the squares of the absolute values of each side of the equation

$$m(1 + 2i) = (1 + i)^r$$

we obtain the equation

$$5m^2 = 2^r$$

which manifestly has no solutions.

Similar considerations applied to the second relation lead to the equation

$$10\, m^2 = 2^s 5^t$$

which has a solution $s=3$, $t=1$, $m=2$. This gives us the relation

$$(3) = 3(1) - (2).$$

We now introduce the formal definition: (n) is called reducible if it can be expressed in the form

$$(n) = \sum f_r \cdot (n_r),$$

where the n_r are positive integers less than n and the f_r are integers (it can be shown that no change occurs if we allow the f_r to be rational); if no such relation exists we call (n) irreducible. Thus (2) is irreducible, while (3) is reducible. We find that (4), (5), (6) are irreducible but that

$$(7) = -(1) + 2(2)$$
$$(8) = 5(1) - (2) - (5).$$

Consideration of these examples suggested the following theorem:

THEOREM. *A condition necessary and sufficient for the reducibility of* (m) *is that the largest prime factor* $l(m)$ *of* $1+m^2$ *should be less than* $2m$.

We can verify this in the early cases quoted:

n	$1+n^2$	$l(n)$		$2n$
2	5	5	>	4
3	10	5	<	6
4	17	17	>	8
5	26	13	>	10
6	37	37	>	12
7	50	5	<	14
8	65	13	<	16
9	82	41	>	18

This theorem can be established constructively by elementary methods: an algorithm for carrying out the reduction of (n) when it is possible can be given, granted that the factorization of $1+r^2$ is known for $r \leqq n$. A listing of the reductions of (n) for $n \leqq 2089$ is available.

The irreducible (n) are analogous to the ordinary prime numbers and it is possible to ask whether there are theorems about them similar to theorems about prime numbers. In the first place there is an analog of Euclid's theorem: there *is* an infinite number of irreducible arctangents. We can also ask whether there is an infinite number of reducible arctangents—this is also true. Both these results are elementary. Gauss conjectured that the number of ordinary prime numbers $<n$ is approximately $n/\log n$ and this "Prime Number Theorem" was proved much later. Observation of the density of the irreducible arctangents suggests that their density is $\log 2 = .6931$—some theoretical evidence in support of this is available, but the result seems very difficult to prove.

It can also be shown that the arctangent of any rational number can be expressed in terms of the irreducible integral arctangents. For instance

$$(100/17) = (6) + (290) - (4836).$$

This work was carried out (in part) under NSF grant 3909 and 11236. The author is indebted for some references to A. A. Albert, D. Estes, I. Kaplansky, G. Pall, J. Robinson.

This is an account of part of a course with the same title given at California Institute of Technology in 1966. It is an extended version of a lecture given at University of California at Santa Barbara and Riverside, at University of Miami and Rice University.

References to 1

1. L. Henkin, Sums of squares, in Summer Institute for symbolic logic, Cornell University 1957, 2nd edition, IDA, (1962) 284–291.

2. I. Kaplansky, Theory of fields, unpublished manuscript, 1970.

3. G. Kreisel, Sums of squares, in Summer Institute for symbolic logic, Cornell University 1957, 2nd edition, IDA, (1962) 313–320.

4. A. Robinson, On ordered fields and definite functions, Math. Ann., 130 (1955) 257–271; 405–409.

5. ———, Introduction to model theory and to the metamathematics of algebra, North Holland, Amsterdam, 1965.

6. J. Robinson, Existential definiability in arithmetic Trans. Amer. Math. Soc. 72, (1952) 437–444.

7. ———, The decision problem for fields, in Symposium on the theory of models, North Holland, Amsterdam, 1965.

8. R. M. Robinson, The undecidability of pure transcendental extensions of real fields, Z. Math. Logik Gunrdlagen Math., 10 (1964) 275–282.

9. A. Tarski and J. C. C. McKinsey, A decision method for elementary algebra and geometry, University of Calif. Press, Berkeley, 1951.

10. O. Taussky, Sums of Squares, Matematička Biblioteka, Beograd, 41 (1969) 19–27.

11. F. van der Blij, History of the octaves, Simon Stevin, 34 (1960/61) 106–125.

References to 2

1. J. P. Ballantine and D. E. Brown, Pythagorean sets of numbers, this MONTHLY, 45 (1938) 298–301.

2. F. J. M. Barning, On Pythagorean and quasi-Pythagorean triangles and a generation process with the help of unimodular matrices, Math. Centrum Amsterdam, 1963.

3. F. Gross, On the equation $f^n+g^n=h^n$, this MONTHLY, 73 (1966) 1093–1096.

4. G. Iyer, On certain functional equations, J. Indian Math. Soc., 3 (1939) 312–315.

5. J. Mariani, The group of the Pythagorean numbers, this MONTHLY, 69 (1962) 125–128.

6. O. Neugebauer and A. Sachs, Mathematical cuneiform texts, New Haven, 1945.

7. N. E. Sexauer, Pythagorean triples over Gaussian domains, this MONTHLY, 73 (1966) 829–834.

8. W. Sierpiński, Pythagorean triangles, Yeshiva University, New York, 1962.

9. O. Taussky, A generalization of the Pythagorean forms, to be published.

10. H. Zassenhaus, What is an angle?, this MONTHLY, 61 (1954) 369–378.

References to 3

1. P. Bateman, Problem E2051, this MONTHLY, 76 (1969) 190–191.

2. E. Bender, Classes of matrices over an integral domain, Illinois J. Math., 11 (1967) 697–702.

3. ———, Classes of matrices and quadratic fields, Linear Algebra and Appl., 1 (1968) 195–201.

4. ———, Characteristic polynomials of symmetric matrices, Pacific J. Math., 25 (1968) 433–441.

5. H. Cohn, Decomposition into four integral squares in the fields of $\sqrt{2}$, $\sqrt{3}$, Amer. J. Math., 82 (1960) 301–322.

6. C. W. Curtis and I. Reiner, Representation theory of finite groups and associative algebras, Interscience, New York, 1962.

7. L. E. Dickson, Integers represented by positive ternary quadratic forms, Bull. Amer. Math. Soc., 33 (1927) 63–70.

8. J. D. Dixon, Another proof of Lagrange's four square theorem, this MONTHLY, 71 (1964) 286–88.

9. J. Dzewas, Quadratsummen in reell-quadratischen Zahlkörpern, Math. Nachr., 21 (1960) 233–284.

10. D. K. Faddeev, On the characteristic equations of rational symmetric matrices, Doklady Akad. Nauk. SSSR, 58 (1947) 753–754.

11. O. Fraser and B. Gordon, On representing a square as the sum of three squares, this MONTHLY, 76 (1969) 922–923.

12. P. R. Halmos, Note on almost-universal forms, Bull. Amer. Math. Soc., 44 (1938) 141–144.

13. H. Heilbronn, On the representation of a rational as a sum of four squares by means of regular functions, J. London Math. Soc., 39 (1964) 72–76.

14. H. Maass, Über die Darstellung totalpositiver Zahlen des Körpers $R(\sqrt{5})$ als Summe von drei Quadraten, Abh. Math. Sem. Univ. Hamburg, 14 (1941) 185–191.

15. ———, Modulformen und quadratische Formen über dem quadratischen Zahlkörper $R(\sqrt{5})$, Math. Ann., 118 (1941) 65–84.

16. M. Newman, Subgroups of the modular group and sums of squares, Amer. J. Math., 82 (1960) 761–778.

17. G. Pall, On sums of squares, this MONTHLY, 40 (1933) 10–18.

18. S. Ramanujan, On the expression of a number in the form $ax^2+by^2+cz^2+du^2$, Proc. Cambridge Phil. Soc., 19 (1917) 11–21.

19. R. A. Rankin, On the representation of a number as the sum of any number of squares, and in particular of twenty, Acta Arith., 7 (1962) 399–407.

20. A. P. Sapiro, Characteristic polynomials of symmetric matrices, Sibirsk Mat. Z., 3 (1962) 280–291.

21. C. L. Siegel, Darstellung total positiver Zahlen durch Quadrate, Math. Z., 11 (1921) 246–275.

22. O. Taussky, On matrix classes corresponding to an ideal and its inverse, Illinois J. Math., 1 (1957) 108–113.

23. ———, Classes of matrices and quadratic fields, Pacific J. Math., 1 (1951) 127–132.

24. ———, Classes of matrices and quadratic fields II, J. London Math. Soc., 27 (1952) 237–239.

25. R. C. Thompson, Problem E1814, this Monthly, 72 (1965) 782; (1967) 200.

References to 4

1. C. S. Ballantine, Products of positive definite matrices III, J. Algebra, 10 (1968) 174–182.

2. ———, A note on the matrix equation $H=AP+PA^*$, Linear Algebra and Appl., 2 (1969) 37–47.

3. ———, Products of positive definite matrices II, to appear.

4. K. Barner, Über die quaternäre Einheitsform in total reellen algebraischen Zahlkörpern, J. Reine Angew. Math., 229 (1968) 194–208.

5. G. Birkhoff, Linear transformations and invariant cones, this MONTHLY, 74 (1967) 274–276.

6. E. Calabi, Linear systems of real quadratic forms, Proc. Amer. Math. Soc., 15 (1964) 844–846.

7. D. H. Carlson, On real eigenvalues of complex matrices, Pacific J. Math., 15 (1965) 1119–1129.

8. P. Erdös and Ch. Ko, On definite quadratic forms, which are not the sum of two definite or semidefinite forms, Acta Arith., 3 (1939) 102–122.

9. D. Estes and G. Pall, The definite octonary quadratic forms of determinant 1, Illinois J. Math., 14(1970) 159–163.

10. M. Fiedler and V. Pták, Some generalizations of positive definiteness and monotonicity, Numer. Math., 9 (1966) 163–172.

11. W. Givens, Fields of values of a matrix, Proc. Amer. Math. Soc., 3 (1952) 206–209.

12. C. Hermite, Lettres de M. Hermite à M. Jacobi, 2nd letter, in Oeuvres 1, Gauthier-Villars, Paris, 1905, 122–135.

13. M. R. Hestenes, Pairs of quadratic forms, Linear Algebra and Appl., 1(1968) 397–407.

14. M. Kneser, Klassenzahlen definiter quadratischer Formen, Archiv Math., 8 (1957) 241–250.

15. ———, unpublished.

16. Ch. Ko, On the representation of a quadratic form as a sum of squares of linear forms, Quart. J. Math. Oxford, 8 (1937) 81–98.

17. ———, Determination of the class number of positive quadratic forms in nine variables with determinant unity, J. London Math. Soc., 13 (1938) 102–110.

18. ———, On the positive definite quadratic forms with determinant unity, Acta Arith., 3 (1939) 75–85.

19. M. Koecher, Positivitätsbereiche im R^n, Amer. J. Math., 53 (1957) 575–596.

20. A. Korkine and G. Zolotareff, Sur les formes quadratiques positives, Math. Ann., 11 (1877) 242–292.

21. A. Lyapunov, Problème général de la stabilité du mouvement, Commun. Soc. Math., Kharkov, (1892), (1893).

22. W. Magnus, Über die Anzahl der in einem Geschlecht enthaltenen Klassen von positiv definiten quadratischen Formen, Math. Ann., 114 (1937) 465–475; Berichtigung, Math. Ann., 115 (1938) 643–644.

23. J. Milnor, in W. H. Greub, Linear Algebra, Academic Press, New York, 1965.

24. H. Minkowski, Mémoire sur la théorie des formes quadratiques à coefficients entiers, Ges. Abh. 1, Chelsea, New York, 1967. 1–144.

25. L. J. Mordell, A new Waring problem with squares of linear forms, Quart. J. Math. Oxford, 4 (1930) 276–280.

26. ———, The definite quadratic forms in eight variables with determinant unity, J. Math. Pure Appl., 17 (1938) 41–46.

27. M. Newman and O. Taussky, On a generalization of the normal basis in abelian algebraic number fields, Commun. Pure Appl. Math., 9 (1956) 85–91.

28. ——— and ———, Classes of definite unimodular circulants, Canadian Math. J., 9 (1956) 71–73.

29. D. G. Quillen, On the representation of hermitian forms as sums of squares, Invent. Math., 5 (1968) 237–242.

30. R. Redheffer, Remarks on a paper of Taussky, J. Algebra, 2 (1965) 42–47.

31. H. Schneider, Positive operators and an inertia theorem, Numer. Math., 7 (1965) 11–15.

32. P. Stein, Some general theorems on iterants, J. Res. Nat. Bur. Standards, 48 (1952) 82–83.

33. ———, On the range of two functions of positive definite matrices, J. Algebra, 2 (1965) 350–353.

34. ——— and A. Pfeffer, On the ranges of two functions II, ICC Bull., 6 (1967) 81–86.

35. O. Taussky, Automorphs and generalized automorphs of quadratic forms treated as characteristic value relations, Linear Algebra and Appl., 1 (1968) 349–356.

36. ———, Unimodular integral circulants, Math. Z., 63 (1955) 286–289.

37. ———, Problem 4846, this MONTHLY, 66 (1959) 427.

38. ———, Positive definite matrices and their role in the study of the characteristic roots of general matrices, Advances in Math., 2 (1968) 175–186.

39. ———, Matrices C with $C^n \to 0$, J. Algebra, 1 (1964) 5–10.

40. ———, Matrix Theory Research Problem, Bull. Amer. Math. Soc., 71 (1965) 711.

41. ———, Positive definite matrices in 'Inequalities', O. Shisha, Ed., Academic Press, 1967.

42. ———, Stable Matrices in 'Programmation en Analyse Numérique,' J. L. Rigal, Ed., Cahiers Centre Math. Rech. Sci. 1968.

43. R. C. Thompson, Unimodular group matrices with rational integers as elements, Pacific J. Math., 14 (1964) 719–726.

44. ———, Classes of definite group matrices, Pacific J. Math., 17 (1966) 175–190.

45. B. L. van der Waerden, Die Reduktionstheorie der positiven quadratischen Formen, Acta Math., 96 (1956) 265–309.

46. E. P. Wigner, On weakly positive matrices, Canadian J. Math., 15 (1965) 313–317.

47. ——— and M. M. Yanase, On the positive semidefinite nature of a certain matrix expression, Canadian J. Math., 16 (1964) 397–406.

48. M. Wonenburger, Simultaneous diagonalization of symmetric bilinear forms, J. Math. Mech., 15 (1966) 617–622.

49. Yik-Hoi Au Yeung, A theorem on a mapping from a sphere to the circle and the simultaneous diagonalization of two hermitian matrices, Proc. Amer. Math. Soc., 20 (1969) 545–548.

50. ———, Some theorems on the real pencil and diagonalization of two hermitian bilinear functions, Proc. Amer. Math. Soc., 23 (1969) 246–254.

References to 5

1. E. Artin and O. Schreier, Algebraische Konstruktion reeller Körper, Abh. Math. Sem. Univ. Hamburg, 5 (1926) 83–115.

2. E. Bender, The dimensions of symmetric matrices with a given minimum polynomial, Linear Algebra and Appl. To appear.

3. J. W. S. Cassels, On the representation of rational functions as sums of squares, Acta Arith., 9 (1964) 79–82.

4. ———, Représentations comme somme de carrés, Les tendances géométriques en algèbre et théorie des nombres, 55–65, Paris, 1966.

5. P. Chowla, On the representation of -1 as a sum of squares in a cyclotomic field, J. Number Theory, 1 (1969) 208–210.

6. B. Fein and B. Gordon, On the representation of -1 as a sum of two squares in an algebraic number field, J. Number Theory, to appear.

7. H. Kneser, Verschwindende Quadratsummen in Körpern, Jber. Deutsch. Math.-Verein, 44 (1934) 143–146.

8. F. Krakowski, Eigenwerte und Minimalpolynome symmetrischer Matrizen in kommutativen Körpern, Comment. Math. Helv., 32 (1958) 224–240.

9. A. Pfister, Zur Darstellung von -1 als Summe von Quadraten in einem Körper, J. London Math. Soc., 40 (1965) 159–165.

10. J.-P. Serre, Extension de corps ordonnes, C. R. Acad. Sci. Paris, 229 (1949) 576–577.

11. J. Smith, The equation $-1=x^2+y^2$ in certain number fields, J. Number Theory, to appear.

12. R. C. Thompson, Commutators in the special and general linear groups, Trans. Amer. Math. Soc., 101 (1961) 16–33.

13. B. L. van der Waerden, Problem, Jber. Deutsch. Math-Verein, 42 (1932) 71.

14. ———, Modern Algebra I, Frederick Ungar, New York, 1966.

References to 6

1. A. A. Albert, Quadratic forms permitting composition, Ann. Math. (2), 43 (1942) 161–177.

2. ———, Quadratic forms permitting composition, unpublished manuscript, 1962.

3. H. Brandt, Der Kompositionsbegriff bei den quaternären quadratischen Formen, Math. Ann., 91 (1924) 300–315.

4. C. Chevalley, The algebraic theory of spinors, Columbia University Press, New York, 1954.

5. C. W. Curtis, The four and eight square problem and division algebras, in Studies in Modern Algebra 2, A. A. Albert, ed., (1963) 100–125.

6. J. Dieudonné, A problem of Hurwitz and Newman, Duke Math. J., 20 (1953) 381–389.

7. B. Eckman, Gruppentheoretischer Beweis der Satzes von Hurwitz-Radon über die Komposition der quadratischen Formen, Comment. Math. Helv., 15 (1942/3), 358–366.

8. H. Freudenthal, Oktaven, Ausnahmegruppen und Oktavengeometrie, Utrecht, 1951.

9. M. Gerstenhaber, On semicommuting matrices, Math. Z., 83 (1964) 250–260.

10. F. Hausdorff, Zur Hilbertschen Lösung des Waringschen Problems, Math. Ann., 67 (1909) 301–305.

11. A. Hurwitz, Über die Komposition der quadratischen Formen, Math. Ann., 88 (1923) 1–25.

12. N. Jacobson, Composition algebras and their automorphisms, Rend. Circ. Mat. Palermo 7, (1958) 55–80.

13. P. Jordan, J. von Neumann and E. Wigner, On an algebraic generalization of the quantum mechanical formalism, Ann. Math., 2 (1934) 29–64.

14. I. Kaplansky, Composition of binary quadratic forms, Studia Math., 31 (1968) 85–92.

15. ———, Infinite-dimensional quadratic forms permitting composition, Proc. Amer. Math. Soc., 4 (1953) 956–96.

16. P. Kustaanheimo and E. Stiefel, Perturbation theory of Kepler motion based on spinor regularization, J. Reine Angew. Math., 218 (1965) 204–219.

17. K. McCrimmon, A proof of Schafer's conjecture for infinite dimensional forms admitting composition, J. Algebra, 5 (1967) 72–83.

18. M. H. A. Newman, Note on an algebraic theorem of Eddington, J. London Math. Soc., 7 (1932) 93–99.

19. A. Pfister, Multiplikative quadratische Formen, Archiv Math., 16 (1965) 363–370.

20. J. Putter, Maximal sets of anti-commuting skew-symmetric matrices, J. London Math. Soc., 42 (1967) 303–308.

21. R. D. Schafer, Forms permitting composition, Advances in Math., 4 (1970) 127–148.

22. W. Scharlau, Quadratische Formen und Galois-Cohomologie, Invent. Math., 4 (1967) 238–264.

23. O. Taussky, A determinantal identity for quaternions and a new eight square identity, J. Math. Anal. Appl., 15 (1966) 162–164.

24. B. Walsh, The scarcity of cross products on euclidean spaces, this MONTHLY, 74 (1967) 188–194.

25. H. Zassenhaus and W. Eichhorn, Herleitung von acht- und sechzehn-Quadrate–Identitäten mit Hilfe von Eigenschaften der verallgemeinerten Quaternionen und der Cayley-Dicksonschen Zahlen, Archiv Math., 17 (1966) 492–496.

References to 7

1. J. F. Adams, Vector fields on spheres, Ann. Math., 75 (1962) 603–632.

2. ———, P. D. Lax and R. S. Phillips, On matrices whose real linear combinations are non-singular, Proc. Amer. Math. Soc., 16 (1965) 318–322; Proc. Amer. Math. Soc., 17 (1966) 945–947.

3. A. A. Albert, Absolute-valued real algebras, Ann. Math., 48 (1947) 495–501.

4. ———, Absolute-valued real algebras, Bull. Amer. Math. Soc., 55 (1949) 763–768.

5. R. Arens, Linear topological division algebras, Bull. Amer. Math. Soc., 53 (1947) 623–630.

6. M. F. Atiyah, The role of algebraic topology in mathematics, J. London Math. Soc., 41 (1966) 63–69.

7. F. A. Behrens, Über Systeme reeller algebraischer Gleichungen, Composito Math., 7 (1939) 1–19.

8. G. Benneton, Sur l'arithmétique des quaternions et des biquaternions, Ann. Sci. École Norm. Sup., (3), 60 (1943) 173–214.

9. R. Bott and J. Milnor, On the parallelizability of the spheres, Bull. Amer. Math. Soc., 64 (1958) 87–89.

10. H. Butts and D. Estes, Modules and binary quadratic forms over integral domains, Linear Algebra and Appl., 1 (1968) 153–180.

11. H. Cohn and G. Pall, Sums of four squares in a quadratic ring, Trans. Amer. Math. Soc., 105 (1962) 536–556.

12. H. S. M. Coxeter, Integral Cayley numbers, Duke Math. J., 13 (1946) 561–578.

13. M. L. Curtis and J. Dugundji, Groups which are cogroups, manuscript.

14. L. E. Dickson, On quaternions and their generalizations and the history of the eight square theorem, Ann. Math., 20 (1919) 155–171.

15. W. Eichhorn, Funktionalgleichungen in Vektorräumen, Kompositionsalgebren und Systeme partieller Differentialgleichungen, Aequationes Math., 2 (1969) 287–303.

16. D. Estes and G. Pall, Modules and rings in the Cayley algebra, J. Number Theory, 1 (1969) 163–178.

17. R. Fueter, Die Theorie der regulären Funktionen einer Quaternionenvariablen, Comptes Rendus du congrès int. des mathématiciens, Oslo, 1936, 75–91.

18. I. Gelfand. Normierte Ringe, Mat. Sbornik N. S. 9 (51) (1941) 3–24.

19. J. W. Givens, Tensor coordinates of linear spaces, Ann. Math., 38 (1937) 355–385.

20. W. R. Hamilton, Mathematical papers, Vol. III, Algebra, Cambridge, 1967.

21. H. Hopf, Ein topologischer Beitrag zur reellen Algebra, Comm. Math. Helv., 13 (1941) 219–239.

22. A. Hurwitz, Über die Zahlentheorie der Quaternionen, Ges. Abh., 2, Birkhäuser, 1933, 303–330.

23. N. Jacobson and O. Taussky, Locally compact rings, Proc. Nat. Acad. Sci. USA, 21 (1935) 106–108.

24. I. Kaplansky, Submodules of quaternion algebras, Proc. London Math. Soc., 19 (1969) 219–232.

25. M. Kervaire, Non-parallelizability of the n-sphere for $n>7$, Proc. Nat. Acad. Sci. USA, 14 (1958) 280–283.

26. H. Kestelman, Anticommuting linear transformations, Canadian J. Math., 13 (1961) 614–624.

27. P. J. C. Lamont, Arithmetics in Cayley's algebra, Proc. Glasgow Math. Assoc., 6 (1963) 99–106.

28. Yu. U. Linnik, Quaternions and Cayley numbers; some applications of the arithmetic of quaternions, Uspehi Mat. Nauk (N.S.) 4 (33) (1949) 49–98.

29. ———, Quaternions and Cayley numbers, Math. Centrum Amsterdam, Rapport ZW-1951-002, 1951.

30. R. Lipschitz, Recherches sur les transformations, par des substitutions réelles d'une somme de deux ou de trois carrés en elle-même, J. Math. Pures Appl., 4, 2 (1886) 373–439.

31. D. Lissner, Outer product rings, Trans. Amer. Math. Soc., 116 (1965) 526–535.

32. K. Mahler, On ideals in the Cayley-Dickson algebra, Proc. Roy. Irish Acad. (A), 48 (1943) 123–133.

33. S. Mazur, Sur les anneaux linéaires, C. R. Acad. Sci. Paris, 207 (1938) 1025–1027.

34. G. Pall, On generalized quaternions, Trans. Amer. Math. Soc., 59 (1946) 280–332.

35. G. Pall and O. Taussky, Applications of quaternions to the representations of binary quadratic form as a sum of four squares, Proc. Roy. Irish Acad. (A), 58 (1957), 23–38.

36. ——— and O. Taussky, Factorization of Cayley numbers, J. Number theory, 2 (1970) 74–90.

37. L. S. Pontryagin, Topological groups; Topological groups, Transl. by A. Brown, Gordon and Breach, New York, 1966.

38. J. Radon, Lineare Schären orthogonaler Matrizen, Abh. Math. Sem. Univ. Hamburg, 1 (1922) 1–14.

39. R. A. Rankin, A certain class of multiplicative functions, Duke Math. J., 13 (1946) 281–306.

40. H. Samelson, Über die Sphären, die als Gruppenräume auftreten, Comm. Math. Helv., 13 (1940/41) 149–155.

41. E. Stiefel, Richtungsfelder und Fernparallelismus in n-dimensionalen Mannigfaltigkeiten, Comm. Math. Helv., 8 (1935/6) 3–51.

42. ———, On Cauchy-Riemann equations in higher dimensions, J. Res. Nat. Bur. Standards, 48 (1952) 395–398.

43. M. H. Stone, On the theorem of Gelfand-Mazur, Ann. Polon. Math., 24 (1952) 238–240.

44. R. G. Swan, Vector bundles and projective modules, Trans. Amer. Math. Soc., 105 (1962) 264–277.

45. O. Taussky, Analytical methods in hypercomplex systems, Compositio Math., 3 (1936) 399–407.

46. ———, An algebraic property of Laplace's differential equation, Quart. J. Math. Oxford, 10 (1939) 99–103.

47. ———, (1, 2, 4, 8)-sums of squares and Hadamard matrices, Proc. Symp. on Combinatorics, Los Angeles, 1968, to appear.

48. L. Tornheim, Normed fields over the real and complex numbers, Michigan Math. J., 1 (1952) 61–68.

49. F. van der Blij and T. A. Springer, The arithmetics of octaves of the group G_2, Nederl. Akad. Wetensch. Proc., 62A (1959) 406–418.

50. F. Wright, Absolute valued algebras, Proc. Nat. Acad. Sci. USA, 39 (1953) 330–332.

51. ———, Absolute valued algebras, Proc. Amer. Math. Soc., 11 (1960) 861–866.

References to 8

1. E. Artin, Über die Zerlegung definiter Funktionen in Quadrate, Abh. Math. Sem. Univ. Hamburg, 5 (1926) 100–115.

2. J. Ax, On ternary definite rational functions, Proc. London Math. Soc., to appear.

3. L. Carlitz, Sums of squares of polynomials, Duke Math. J., 3 (1937), 1–7.

4. D. W. Dubois, Note on Artin's solution of Hilbert's 17th problem, Bull. Amer. Math. Soc., 73 (1967) 540–541.

5. W. Habicht, Über die Zerlegung strikt definiter Formen in Quadrate, Comment. Math. Helv., 12 (1946) 317–322.

6. ———, Zerlegung strikte definiter Formen in Quadrate, Comment. Math. Helv., 12 (1940) 317–322.

7. D. Hilbert, Über die Darstellung definiter Formen als Summe von Formenquadraten, Ges. Abh. 2, Springer, Berlin, 1933, 154–161.

8. ———, Mathematische Probleme, in particular problem 17, Ges. Abhandlungen III, Springer, Berlin, 1935, 290–329.

9. E. Landau, Über die Darstellung definiter Funktionen als Summe von Quadraten, Math. Ann., 62 (1906) 272–285.

10. T. S. Motzkin, Algebraic inequalities, in Inequalities, Ed. O. Shisha, Academic Press, New York, 1967, 199–203.

11. A. Pfister, Zur Darstellung definiter Funktionen als Summe von Quadraten, Invent. Math., 4 (1967), 229–237.

12. R. M. Robinson, Some definite polynomials which are not sums of squares of real polynomials (abstract), Notices, Amer. Math. Soc., 16 (1969) 554.

References to 9

1. Gösta Bucht, Arkiv. Math. Astron. and Physik. 6, No. 30.

2. H. Reichardt, Über Normalkörper mit Quaternionengruppe, Math. Z., 41 (1936) 218–221.

3. B. L. van der Waerden, Problem, Jber. Deutsch. Math.-Verein., 43 (1933) 61.

References to 10

1. S. D. Chowla and John Todd, The density of reducible integers, Canadian J. Math., 1 (1949) 297–299.

2. D. H. Lehmer, On arccotangent relations for π, this MONTHLY, 45 (1938) 657–664.

3. John Todd, A problem on arctangent relations, this MONTHLY, 56 (1949) 517–528.

4. ———, Table of arctangents of rational numbers, National Bureau of Standards, Applied Math. Series, No. 11 (1951–1965), U. S. Government Printing Office, Washington, D. C.

This list is not intended to be complete in any respect.

HARDY'S "A MATHEMATICIAN'S APOLOGY"

L. J. MORDELL, St. Johns College, Cambridge, England

A reprint of this most interesting book appeared in 1967 with a foreword by C. P. Snow, Hardy's friend of long standing.

It has often been reviewed and highly praised, but there are, however, some opinions expressed by Hardy which, perhaps, have not been adequately dealt with by other reviewers. Furthermore, Snow's foreword calls for some comment, especially his references to Ramanujan (1887–1920). He writes that after Hardy and Littlewood read the manuscript sent by Ramanujan to Hardy (probably Jan. 16, 1913), "they knew, and knew for certain" that he "was a man of genius .... It was only later that Hardy decided that Ramanujan was, in terms of *natural* mathematical genius, in the class of Gauss and Euler; but that he could not expect, because of the defects of his education, and because he had come on the scene too late in the line of mathematical history, to make a contribution on the same scale." While one would readily accept that Ramanujan was a man of genius, the comparison with Gauss and Euler is very farfetched. I have some difficulty in believing that Hardy made such a statement, or at any rate made it in this form. What does natural mathematical genius mean? Undoubtedly Ramanujan was outstanding in some aspects of mathematics and had great potentialities. But this is not enough. What really matters is what he did, and one cannot accept such a comparison with Euler and Gauss, whose many-sided contributions were of fundamental importance and changed the face of mathematics. In fact in 1940, in the book on Ramanujan, Hardy said "I cannot imagine anybody saying with any confidence, even now, just how great a mathematician he was and still less how great a mathematician he might have been."

Snow says that Ramanujan, as is commonly believed, was the first Indian to be elected (2 May, 1918) a fellow of the Royal Society. He was the second. The first was Ardaseer Cursetjee (1808–1877), shipbuilder and engineer, F.R.S. 27 May, 1841. Snow notes that Ramanujan was elected a fellow of Trinity four years after his arrival in England and continues, "it was a triumph of academic uprightness that they should have elected Hardy's protégé Ramanujan at a time when Hardy was only just on speaking terms with some of the electors and not at all with others." It is well that the merits of a fellowship candidate are judged by the quality of his original work and not by the political views of his sponsors.

Let us examine some of the views expressed by Hardy. They are sometimes stated too categorically, regardless of exceptions and limitations. A number of them had their origin in what he says, most gloomily, in the very first section of the Apology: "It is a melancholy experience for a professional mathematician to find himself writing about mathematics. The function of a mathematician is to do something, to prove new theorems, to add to mathematics, and not to talk about what he or other mathematicians have done."

His practice many years ago does not conform with this statement. He recalls

in Section 6 that he did talk about mathematics in his 1920 Oxford inaugural lecture, which actually contains an apology for mathematics. Further in 1921, he gave an address on Goldbach's theorem to the Mathematical Society of Copenhagen. In this, he did talk about what he and other mathematicians had done. Such talks render a real service to mathematics and many have found great pleasure and inspiration in listening to or reading such expositions. Hardy had followed the practice of many eminent mathematicians in giving them. These have contributed to the richness and vividness of mathematics and make it a living entity. Without them, mathematics would be much the poorer.

No mathematician can always be producing new results. There must inevitably be fallow periods during which he may study and perhaps gather ideas and energy for new work. In the interval, there is no reason why he should not occupy himself with various aspects of mathematical activity, and every reason why he should. The real function of a mathematician is the advancement of mathematics. Undoubtedly the production of new results is the most important thing he can do, but there are many other activities which he can initiate or participate in. Hardy had his full share of these. He took a leading part in the reform of the mathematical tripos some sixty years ago. Before then, it was looked upon as a sporting event, reminding one of the Derby, and was out of touch with continental mathematics. A mathematician can engage in the many administrative aspects of mathematics. Hardy was twice secretary and president of the London Mathematical Society and, while so occupied, must have done an enormous amount of unproductive work. He served on many committees dealing with mathematics and mathematicians. He wrote a great many obituary notices. He was well aware that a professor of mathematics is a representative of his subject in his University. This entails many duties which cannot be called doing mathematics.

His reference to a melancholy experience shows how much he took to heart and suffered from the loss of his creative powers. The result is, as Snow says, that the *Apology* is a book of haunting sadness.

Further in this first section, he says despairingly, "if then I find myself writing not mathematics but 'about' mathematics, it is a confession of weakness, for which I may rightly be scorned or pitied by younger and more vigorous mathematicians. I write about mathematics because, like any other mathematician who has passed sixty, I have no longer the freshness of mind, the energy, or the patience to carry on effectively with my proper job." He had been for many years a most active mathematician and his collected works now being published will consist of seven volumes. It seems almost nonsense to say that anyone would scorn or pity him, and the use of the term 'rightly' is even more nonsensical.

We all know only too well that with advancing age we are no longer in our prime, and that our powers are dimmed and are not what they once were. Most of us, but not Hardy, accept the inevitable. There are still many consolations. We can perhaps find pleasure in thinking about some of our past work. We can

read what others are doing, but this may not be easy since many new techniques have been evolved, sometimes completely changing the exposition of classical mathematics. Various reviews, however, may give one some idea of what has been done. (We can still be of service to younger mathematicians.)

His statement about a mathematician who has passed sixty is far too sweeping and any number of instances to the contrary can be mentioned, even among much older people. One need only note some recent Cambridge and Oxford professors. Great activity among octogenarians is shown by Littlewood, his lifelong collaborator, Sydney Chapman, his former pupil and collaborator, and myself. There is also Besicovitch in the seventies. Davenport, who had passed sixty, was as active and creative as ever, and his recent death is a very great loss to mathematics since he could have been expected to continue to produce beautiful and important work.

The question of age was ever present in Hardy's mind. In Section 4, he says, "No mathematician should ever allow himself to forget that mathematics, more than any other art or science, is a young man's game." It seems that he could not reconcile himself to growing old. For further on, he ways, "I do not know an instance of a major mathematical advance initiated by a man past fifty." This may be so, but much depends on the definition of the *advance*. But there is no need to be troubled about it. Much important work has been done by men after the age of fifty.

A number of Hardy's statements must be qualified. In Section 2, he says, that "good work is not done by 'humble' men. It is one of the first duties of a professor, for example, in any subject, to exaggerate a little both the importance of his subject and his own importance in it. A man who is always asking, 'Is what I do worthwhile?' and 'Am I the right person to do it?' will always be ineffective himself and a discouragement to others."

Though one may naturally have a better opinion of one's work than others have, there are many exceptions to his statement. I never knew Davenport to exaggerate or emphasize the importance of his work, but he was a most effective mathematician and a very successful supervisor of research. Prof. Frechet told me a few years ago, that when Norbert Wiener was working with him a long time ago, he was always asking, "Is my work worthwhile?", "Am I slipping?", etc. S. Chowla is as modest and humble a mathematician as I know of, but he inspires many research students.

We comment on some more of Hardy's statements about mathematics. One of the most surprising is in Section 29, "I do not remember having felt, as a boy, any *passion* for mathematics, and such notions as I may have had of the career of a mathematician were far from noble. I thought of mathematics in terms of examinations and scholarships; I wanted to beat other boys, and this seemed to me to be the way in which I could do so most decisively."

It has often been said that mathematicians are born and not made. Most great mathematicians developed their keenness for mathematics in their school days. Their ability revealed itself by comparison with the performances of their

schoolmates. Their ambition was to continue the study of mathematics and to take up a mathematical career. Probably no other motive played any part in the decision of most of them.

In Section 3, he considers the case of a man who sets out to justify his existence and his activities. I see no need for justification any more than a poet or painter or sculptor does. As Trevelyan says, disinterested intellectual curiosity is the life blood of real civilization. It is curiosity that makes a mathematician tick. When Fourier reproached Jacobi for trifling with pure mathematics, Jacobi replied that a scientist of Fourier's calibre should know that the end of mathematics is the great glory of the human mind. Most mathematicians do mathematics for the very good reason that they like and enjoy doing it. Davenport told me that he found it "exciting" to do mathematics.

Hardy says that the justifier has to distinguish two different questions. The first is whether the work which he does is worth doing and the second is why he does it, whatever its value may be. He says to the first question: The answer of most people, if they are honest, will usually take one or the other of two forms; and the second form is merely a humbler version of the first, which we need to consider seriously. "I do what I do because it is the one and only thing I can do at all well." It suffices to say that the mathematician felt no need to do anything else.

Hardy is very appreciative of the beauty and aesthetic appeal of mathematics. "A mathematician," he says in Section 10, "like a painter or poet, is a maker of patterns . . . ," and these " . . . must be beautiful. The ideas . . . must fit together in a harmonious way. Beauty is the first test: there is no permanent place in the world for ugly mathematics . . . . It may be very hard to *define* mathematical beauty . . . " but one can recognize it. He discusses the aesthetic appeal of theorems by Pythagoras on the irrationality of $\sqrt{2}$, and Euclid on the existence of an infinity of prime numbers. He says in Section 18, "there is a very high degree of *unexpectedness*, combined with *inevitability* and *economy*. The arguments take so odd and surprising a form; the weapons used seem so childishly simple when compared with the far-reaching results."

I might suggest among other attributes of beauty, first of all, simplicity of enunciation. The meaning of the result and its significance should be grasped immediately by the reader, and these in themselves may make one think, what a pretty result this is. It is, however, the proof which counts. This should preferably be short, involve little detail and a minimum of calculations. It leaves the reader impressed with a sense of elegance and wondering how it is possible that so much can be done with so little.

Somehow, I do not think that Hardy's work is characterized by beauty. It is distinguished more by his insight, his generality, and the power he displays in carrying out his ideas. Many of the results that he obtains are very important indeed, but the proofs are often long and require concentrated attention, and this may blunt one's feelings even if the ideas are beautiful.

Hardy does not define ugly mathematics. Among such, I would mention

those involving considerable calculations to produce results of no particular interest or importance; those involving such a multiplicity of variables, constants, and indices, upper, lower, right, and left, making it very difficult to gather the import of the result; and undue generalization apparently for its own sake and producing results with little novelty. I might also mention work which places a heavy burden on the reader in the way of comprehension and verification unless the results are of great importance.

Hardy had previously said that he could "quote any number of fine theorems from the theory of numbers whose meaning anyone can understand, but whose proofs, though not difficult, may be found tedious." It often happens that there are significant results apparently of some depth, the proof of which can be grasped by those with a minimum of mathematical knowledge. Perhaps I may be pardoned if I give one of my own. The theorem of Pythagoras suggests the problem of finding the integer solutions of the equation $x^2+y^2=z^2$. This was done some 1000 years ago and is not difficult. But suppose we consider the more general equation $ax^2+by^2=cz^2$. This is a real problem in the theory of numbers. Legendre at the end of the eighteenth century gave necessary and sufficient conditions for its solvability. Then when the equation is taken in the normal form, i.e. abc is square-free and $a>0$, $b>0$, $c>0$, Holzer showed in 1953 that a solution existed with $|z|<\sqrt{ab}$, from which it follows that $|x|\leqq\sqrt{bc}$, $|y|\leqq\sqrt{ca}$. I recently found a proof of this result that no one would call tedious by showing that if a solution (x_1, y_1, z_1) existed with $|z_1|>\sqrt{ab}$, then there was another with $|z_2|<|z_1|$. This arose by taking an appropriate line through the point (x_1, y_1, z_1) to meet the conic $ax^2+by^2=cz^2$ in the point (x_2, y_2, z_2). I call this a schoolboy proof, because the only advanced result required is that the equation $lx+my=n$ has an integer solution if l and m are co-prime. A proof of the theorem could have been found by a schoolboy.

We conclude by examining Hardy's views about the utility or usefulness of mathematics. He seems to denigrate the usefulness of 'real' mathematics. In Section 21, he says, "The 'real' mathematics of the 'real' mathematicians, the mathematics of Fermat and Euler and Gauss and Abel and Riemann is almost wholly 'useless.' " This statement is easily refuted. A ton of ore contains an almost infinitesimal amount of gold, yet its extraction proves worthwhile. So if only a microscopic part of pure mathematics proves useful, its production would be justified. Any number of instances of this come to mind, starting with the investigation of the properties of the conic sections by the Greeks and their application many years later to the orbits of the planets. Gauss' investigations in number theory led him to the study of complex numbers. This is the beginning of abstract algebra, which has proved so useful for theoretical physics and applied mathematics. Riemann's work on differential geometry proved of invaluable service to Einstein for his relativity theory. Fourier's work on Fourier series has been most useful in physical investigations. Finally one of the most useful and striking applications of pure mathematics is to wireless telegraphy which had its origin in Maxwell's solution of a differential equation. Many new

disciplines are making use of more and more pure mathematics, e.g., the biological sciences, economics, game theory, and communication theory, which requires the solution of some difficult Diophantine equations. It has been truly said that advances in science are most rapid when their problems are expressed in mathematical form. These in time may lead to advances in pure mathematics.

These remarks may serve as a reply to Hardy's statement that the great bulk of higher mathematics is useless.

It is suggested that one purpose mathematics may serve in war is that a mathematician may find in mathematics an incomparable anodyne. Bertrand Russell says that in mathematics, "one at least of our nobler impulses can best escape from the dreary exile of the actual world." Hardy's comment on this reveals his depressed spirits. "It is a pity," he says, "that it should be necessary to make one very serious reservation—he must not be too old. Mathematics is not a contemplative but a creative subject; no one can draw much consolation from it when he has lost the power or desire to create; and that is apt to happen to mathematicians rather soon." What does he mean when he says mathematics is not a contemplative subject? Many people can derive a great deal of pleasure from the contemplation of mathematics, e.g., from the beauty of its proofs, the importance of its results, and the history of its development. But alas, apparently not Hardy.

A mathematician's apology by G. H. Hardy with a foreword by C. P. Snow, University Press, Cambridge 1967 (18 s.). A paperback edition is available at 8s.

WINNING STRATEGIES FOR THE IDEAL GAME

C. W. HENSON, Duke University

If R is a commutative ring (with identity) then the *ideal game on* R—or simply the R-*game*—is a two-person game in which the players alternate in choosing elements from R, with each choice being subject to certain conditions on the ideal generated by the set of previously chosen elements. Specifically, the player making the nth move must choose $\alpha_n \in R$ so that

(i) α_n is not in the ideal $(\alpha_1, \cdots, \alpha_{n-1})$ generated by the previously chosen elements, and

(ii) $(\alpha_1, \cdots, \alpha_n)$ is a proper ideal. The player who is able to choose α_n so that $(\alpha_1, \cdots, \alpha_n)$ is a maximal ideal is the winner. (If R is a field, so that no first choice is possible, we say that the second player wins.)

If R is Noetherian, so that every increasing chain of ideals has finite length, then every play of the R-game must have a winner. In such a case there is necessarily a winning strategy for one of the players. In specific cases, however, it is difficult to tell which player has the strategy or to give a description of it. In this paper we show that if R is the ring $P[x]$ of polynomials over a PID (principal ideal domain) P, then it is the first player who has the winning strategy.

In particular this is true when R is $Z[x]$ or $F[x, y]$, where Z is the ring of rational integers and F an arbitrary field. The proof uses only elementary facts from the theory of rings and is constructive, in the sense that an explicit description of the strategy can be extracted from the argument. This result follows from the:

MAIN THEOREM. *Let D be an integral domain and $a \neq 0$ an element of D such that $D/(a)$ is a PID but not a field. Then the first player has a winning strategy in the D-game, beginning with the choice of a^2.*

COROLLARY. *Let D be $P[x]$, where P is a PID but not a field. If a is either an irreducible element of P or of the form $x+b(b\in P)$ then the first player has a winning strategy in the D-game, beginning with the choice of a^2.*

Proof. In the first case $D/(a)\simeq(P/(a))[x]$ and in the second $D/(a)\simeq P$. In either case the Main Theorem applies.

Note that the Main Theorem yields a first-player winning strategy in the $P[x]$-game whenever P is an integral domain with a principal maximal ideal.

For the remainder of this paper we fix D and a as in the Main Theorem, and let R be $D/(a^2)$. Also let $\sigma: D\to R$ and $\pi: R\to R/(\sigma a)$ be the canonical homomorphisms and set $\alpha=\sigma(a)$. Then $R/(\alpha)\simeq D/(a)$, so that $R/(\alpha)$ is a PID but not a field. Moreover, R and α satisfy the property

(A) $$\text{For any } \beta\in R,\ \beta\alpha=0\leftrightarrow\beta\in(\alpha).$$

Indeed if $\beta\alpha=0$ then $\sigma(b)=\beta$ and $ba\in(a^2)$ for some $b\in D$. Since D is an integral domain this implies $b\in(a)$, and hence $\beta\in(\alpha)$. The converse follows from the fact that $\alpha^2=\sigma(a^2)=0$.

The rest of this paper is devoted to proving that the second player has a winning strategy in the R-game (Theorem 1) and hence to proving the Main Theorem.

LEMMA 1. *If I is an ideal of R and $b\in I$ satisfies $\pi(I)=(\pi(b))$ in $R/(\alpha)$, then $I=(b)+I\cap(\alpha)$.*

Proof. We must show $I\subseteq(b)+I\cap(\alpha)$. If $a\in I$ then $\pi(a)\in(\pi(b))$ so that $a=cb+d\alpha$, for some $c, d\in R$. But $a-cb\in I$ so that $a\in(b)+I\cap(\alpha)$, as desired.

LEMMA 2. *Let I be an ideal of R. I is prime $\leftrightarrow I=(\alpha)$ or I is maximal. I is maximal $\leftrightarrow I=(\alpha, \beta)$ where $\pi(\beta)$ is irreducible in $R/(\alpha)$.*

Proof. Since $\alpha^2=0$ every prime (thus every maximal) ideal contains (α). In that case I and $\pi(I)$ are associated ideals. Both statements then follow from the fact that $R/(\alpha)$ is a PID.

LEMMA 3. *Let I be an ideal of R;*

(i) *if $I\subseteq(\alpha)$ then I is principal,*

(ii) *there exist $b, c\in R$ so that $I=(b, c\alpha)$. In fact this is true whenever $b\in I$, $\pi(I)=(\pi(b))$ and $I\cap(\alpha)=(c\alpha)$.*

Proof. (i) Let J be the ideal $I:(\alpha)$ so that $J=\{x \mid x\alpha\in I\}$. Then $I\cap(\alpha)=J\alpha$. Choose $c\in J$ so that $\pi(J)=(\pi(c))$, and hence $J=(c)+J\cap(\alpha)$, by Lemma 1. Then $I=I\cap(\alpha)=J\alpha=(c\alpha)$, so that I is principal.

(ii) This follows from (i) using Lemma 1.

We shall call any representation of an ideal I as in Lemma 3 (ii) a *standard representation* of I.

REMARK. Since every ideal of R is finitely generated, R is Noetherian. Thus each proper ideal I has a normal decomposition into primary ideals (cf. [1]). If $I\not\subseteq(\alpha)$ then the prime ideals containing I are all maximal, by Lemma 2. Hence if the normal decomposition of such an I is $I=I_1\cap \cdots \cap I_n$ (where each I_j is an M_j-primary ideal, M_j a maximal ideal) then we have: (i) $M_1, \cdots, M_n$ are the maximal ideals containing I, and (ii) each I_j is the smallest M_j-primary ideal containing I. Also a proper ideal J is M-primary (M a maximal ideal) exactly when $J\supseteq M^n$ for some $n>0$.

If I is a proper ideal not contained in (α), M is a maximal ideal and $I\subseteq M$ then we let I_M be the M-primary component of I. It follows that for any M-primary ideal J, $I+J$ is also M-primary and $I+J=I_M+J$.

LEMMA 4. *If $I=(b, c\alpha)$ in R and $b\notin(\alpha)$ then $I\cap(\alpha)=(b\alpha, c\alpha)$.*

Proof. We must show $I\cap(\alpha)\subseteq(b\alpha, c\alpha)$. If $a=a'\alpha$ is in $I\cap(\alpha)$ then $a=a'\alpha=b'b+c'c\alpha$ for some $b', c'\in R$. Then $b'\in(\alpha)$ since (α) is a prime ideal. That is, $b'=b''\alpha$ for some b'' and $a=b''(b\alpha)+c'(c\alpha)$ as desired.

LEMMA 5. *Let $a\in R$; a is a unit $\leftrightarrow \pi(a)$ is a unit in $R/(\alpha)$.*

Proof. Obviously $\pi(a)$ is a unit if a is one. If $\pi(a)\pi(b)=1$ in $R/(\alpha)$ then $ab=1+c\alpha$ for some $c\in R$. But then $a[b(1-c\alpha)]=1-(c\alpha)^2=1$, so a is a unit.

LEMMA 6. *Let $M=(\alpha, \beta)$ be a maximal ideal of R and let I be an M-primary ideal. Then there exist $k\geqq r\geqq 0$, $k>0$ and $a\in R$ so that $I=(\beta^k+a\alpha, \beta^r\alpha)$ is a standard representation of I.*

Proof. Since I is M-primary there is an $n>0$ so that $I\supseteq M^n$ and hence $\pi(I)\supseteq(\pi(\beta)^n)$. Thus there is a $k>0$ so that $\pi(I)=(\pi(\beta)^k)$. By Lemma 3 there exist $a, c\in R$ so that $I=(\beta^k+a\alpha, c\alpha)$ is a standard representation of I. Thus $\beta^k\alpha\in(c\alpha)$ and there exists $d\in R$ with $(\beta^k-dc)\alpha=0$. This implies that $\pi(\beta)^k\in(\pi(c))$, by property A, and hence there is a unit $\pi(u)$ in $R/(\alpha)$ and $k\geqq r\geqq 0$ such that $\pi(c)=\pi(u)\pi(\beta)^r$. Then $c\alpha=u\beta^r\alpha$ so that $(c\alpha)=(\beta^r\alpha)$ by Lemma 5, completing the proof.

If $M=(\alpha, \beta)$ is a maximal ideal of R and $n>0$, then M^n is generated by the set of elements of the form $\beta^r\alpha^s$, where $r, s\geqq 0$ and $r+s=n$. That is, $M^n=(\beta^n, \beta^{n-1}\alpha)$. Thus if $I=(\beta^k+a\alpha)$ where $k>0$ then I is M-primary, since it contains $(\beta^{2k}, \beta^{2k-1}\alpha)$, and by Lemma 6 every principal M-primary ideal can be expressed in this form. But this implies that $M^n (n>0)$ is not a principal ideal: otherwise $M^n=(\beta^n+a\alpha)$ and $(\beta^n\alpha)=(\beta^{n-1}\alpha)$, using Lemma 4. This implies that $\pi(\beta)^n\in(\pi(\beta)^{n-1})$, by property A, contradicting the irreducibility of $\pi(\beta)$.

DEFINITION. An ideal I of R is *special* $\leftrightarrow$ there is a maximal ideal M so that $I=M^n$, for some $n>0$.

We shall show that the special ideals of R are exactly those proper ideals I of R such that there is a winning strategy for the second player in the R/I-game.

LEMMA 7. *Let M be a maximal ideal of R and I an M-primary ideal. Then there is a $b\in R$ such that (b) is M-primary and $I+(b)$ is special.*

Proof. Let $M=(\alpha, \beta)$ and $I=(\beta^k+a\alpha, \beta^r\alpha)$ as in Lemma 6, $(k\geqq r\geqq 0,\ k>0)$.

Case 1. Assume $a\in(\alpha, \beta^r)$, and hence $I=(\beta^k, \beta^r\alpha)$. If $k>r$ then set $b=\beta^{r+1}$, so that $I+(b)=(\beta^{r+1}, \beta^r\alpha)$. If $k=r$ we set $b=\beta^r+\beta^{r-1}\alpha$ and have $I+(b)=(\beta^r, \beta^{r-1}\alpha)$.

Case 2. Assume $a\notin(\alpha, \beta^r)$ so we may let s be the largest integer such that $a\in(\alpha, \beta^s)$. Thus $0\leqq s<r$, and we may write $a=a'\beta^s+a''\alpha$ where $a'\notin(\alpha, \beta)$. That is, $a\alpha=a'\beta^s\alpha$. Set $b=\beta^{s+1}$ so that $I+(b)=(\beta^{s+1}, a'\beta^s\alpha)$. Since $a'\notin(\alpha, \beta)$ which is maximal, we may write $1=ca'+d\beta+e\alpha$. That is

$$\beta^s\alpha = c(a'\beta^s\alpha) + d\alpha\beta^{s+1}$$

and hence $I+(b)=(\beta^{s+1}, \beta^s\alpha)$. The proof is completed by noting that each (b) is M-primary and each $I+(b)$ is special.

LEMMA 8. *Let (b) be a proper ideal of R which is not contained in (α). Then the primary components of (b) are principal.*

Proof. Let $M=(\alpha, \beta)$ be a maximal ideal containing (b). There is a largest integer k such that $\pi(b)\in(\pi(\beta)^k)$, and $k>0$. Thus we may write $b=a\beta^k+a'\alpha$ where $a\notin(\alpha, \beta)$. Hence there exist c, d, $e\in R$ with $1=ca+d\beta^k+e\alpha$. This leads to $b=(a+da'\alpha)(\beta^k+ca'\alpha)$, so we have expressed b in the form $b_1(\beta^k+b_2\alpha)$ where $b_1\notin(\alpha, \beta)$. This implies that any M-primary ideal containing (b) also contains $(\beta^k+b_2\alpha)$, which is M-primary itself. That is, $(b)_M=(\beta^k+b_2\alpha)$ and the proof is complete.

LEMMA 9. *Let M be a maximal ideal of R and $n>k>0$. Then there is no $b\in R$ such that $M^n+(b)=M^k$.*

Proof. Otherwise, let $M=(\alpha, \beta)$ be maximal and assume $M^n+(b)=M^k$ where $n>k>0$. Passing to $R/(\alpha)$ we see that $b\notin(\alpha)$. Using Lemma 8 we let $(b)_M=(\beta^r+a\alpha)$. Also $M^k=M^n+(b)=M^n+(b)_M$, so that $k=r$ and $M^k=(\beta^n, \beta^k+a\alpha)$. Thus $M^k=(\beta^k, a\alpha)$. By Lemma 4 this implies that

$$(\beta^{k-1}\alpha) = M^k \cap (\alpha) = (\beta^k\alpha, a\alpha)$$

so we may write $a\alpha=a'\beta^{k-1}\alpha$. But this implies that $\beta^n=\beta^{n-k}(\beta^k+a\alpha)-a'\beta^{n-1}\alpha\in(\beta^k+a\alpha)$ and hence that $M^n\subseteq(\beta^k+a\alpha)$. That is, $M^k=(\beta^k+a\alpha)$ which is a contradiction.

LEMMA 10. *Let I be a proper ideal of R. Then there exists $b\in R$ such that $I+(b)$ is special.*

Proof. If I is not contained in (α) and M is any maximal ideal containing I, there is an M-primary ideal (b) so that $I_M+(b)$ is special, by Lemma 7. Thus $I+(b)=I_M+(b)$ and b is as desired.

If $I=(\alpha)$ let b be such that (α, b) is a maximal ideal. If $I\subseteq(\alpha)$ but $I\neq(\alpha)$, then by Lemma 3 we may write $I=(c\alpha)$ where (c) is a proper ideal and $c\notin(\alpha)$. Let $M=(\alpha, \beta)$ be a maximal ideal containing (c) so that $(c)_M=(\beta^k+a\alpha)$ for some $k>0$. Set $b=\beta^{k+1}$; by Lemma 4 $(I+(b))\cap(\alpha)=(b\alpha, c\alpha)$. But $(b\alpha, c\alpha)=((b)+(c)_M)\alpha$ since (b) is M-primary, so

$$(I + (b)) \cap (\alpha) = (\beta^{k+1}, \beta^k + a\alpha)\alpha = (\beta^k\alpha).$$

Hence $I+(b)=(\beta^{k+1}, \beta^k\alpha)$. In each case $I+(b)$ is special and therefore the proof is complete.

THEOREM 1. *The second player has a winning strategy in the R-game. In fact, if I is a proper ideal of R, then the second player has a winning strategy in the R/I-game if I is special, and the first player has a winning strategy in the R/I-game if I is not special.*

Proof. Let I be a special ideal. If $I\neq I+(a)\neq R$ then $I+(a)$ is not special, by Lemma 9. Then there exists $b\in R$ so that $I+(a, b)$ is special, by Lemma 10. Since the maximal ideals are special (and R is Noetherian) it follows that the second player has a winning strategy in the R/I-game whenever I is special. If I is a proper ideal but is not special, then there is a $b\in R$ such that $I+(b)$ is special; that is, the R/I-game has a first-player winning strategy in this case.

Finally, since no principal ideal of R is special there must be a winning strategy for the second player in the R-game.

References

1. D. G. Northcott, Ideal Theory, Cambridge University Press, Cambridge, 1953.
2. Seth Warner, Modern Algebra, Vol. I, Prentice-Hall, Englewood Cliffs, N. J., 1965.

SUMMATION OF THE SERIES $1^n+2^n+\cdots+x^n$ USING ELEMENTARY CALCULUS

L. S. LEVY, University of Wisconsin

It is well known that the sum of the series $1+2+\cdots+x$ is $x(x+1)/2$, a polynomial which we shall call $P_1(x)$; and it is almost as well known that $1^2+2^2+\cdots+x^2=x(x+1)(2x+1)/6=P_2(x)$. More generally, an old theorem (one proof of which is suggested by problem 3, Section 4) states:

THEOREM 1. *For each positive integer n there is exactly one polynomial $P_n(x)$ in x such that*

$$(1) \qquad P_n(x) = 1^n + 2^n + \cdots + x^n \quad \textit{whenever } x \textit{ is a positive integer.}$$

The degree of $P_n(x)$ *is* $n+1$.

The first few of these polynomials are:

$$P_1(x) = \frac{x^2}{2} + \frac{x}{2} = \frac{1}{2}x(x+1),$$

$$P_2(x) = \frac{x^3}{3} + \frac{x^2}{2} + \frac{x}{6} = \frac{1}{6}x(x+1)(2x+1),$$

$$P_3(x) = \frac{x^4}{4} + \frac{x^3}{2} + \frac{x^2}{4} = \frac{1}{4}x^2(x+1)^2,$$

$$P_4(x) = \frac{x^5}{5} + \frac{x^4}{2} + \frac{x^3}{3} - \frac{x}{30} = \frac{1}{30}x(x+1)(2x+1)(3x^2+3x-1),$$

$$P_5(x) = \frac{x^6}{6} + \frac{x^5}{2} + \frac{5x^4}{12} - \frac{x^2}{12} = \frac{1}{12}x^2(x+1)^2(2x^2+2x-1),$$

$$P_6(x) = \frac{x^7}{7} + \frac{x^6}{2} + \frac{x^5}{2} - \frac{x^3}{6} + \frac{x}{42} = \frac{1}{42}x(x+1)(2x+1)(3x^4+6x^3-3x+1),$$

$$P_7(x) = \frac{x^8}{8} + \frac{x^7}{2} + \frac{7x^6}{12} - \frac{7x^4}{24} + \frac{x^2}{12} = \frac{1}{24}x^2(x+1)^2(3x^4+6x^3-x^2-4x+2).$$

The most obvious question one can ask about these polynomials is how to find them. Some more subtle questions are suggested by the following observations from the table:

OBSERVATION 1. $P_n(x)$ has the following factors: $x(x+1)$ for all n, $x(x+1)\cdot(2x+1)$ for all even n; $x^2(x+1)^2$ for all odd $n<1$.

OBSERVATION 2. The coefficient, in $P_n(x)$, of x^{n+1} is always $1/(n+1)$ and that of x^n is always $1/2$.

OBSERVATION 3. With the exception of the term $x^n/2$, all of the terms of $P_n(x)$ have even degree, or else all have odd degree.

The key to all these questions will be an integration formula

$$P_n(x) = n\int_0^x P_{n-1}(t)dt + C_n x \quad \text{for some constant } C_n. \tag{2}$$

To see how (2) can be used, start with $P_1(x) = x^2/2 + x/2$. Then

$$P_2(x) = 2\int_0^x (t^2/2 + t/2)dt + C_2x = x^3/3 + x^2/2 + C_2x.$$

To find C_2, substitute the value $x=1$ into this equation, using the fact that $P_2(1)$ is the sum of *one* term of the series $1^2+2^2+\cdots$, that is, $1=P_2(1)=(1/3)+(1/2)+C_2$. This shows that $C_2=1/6$ as indicated in the table. We will prove, in section 3, that $C_n=0$ whenever n is *odd* and >1. Thus from (2) we get

$$P_3(x) = 3\int_0^x (t^3/3 + t^2/2 + t/6)dt = (x^4/4) + (x^3/2) + (x^2/4) \quad \text{and so on.}$$

Note that the formula $1^3+2^3+\cdots+x^3=(x^4/4)+(x^3/2)+(x^2/4)$ which we obtained from the integration formula (2) is a statement about the sum of a series of *integers*. The main point of this paper is to show (without the complications one encounters in analytic number theory) how properties of continuous functions can be used to obtain properties of the integers.

1. A Pitfall. Before applying calculus to our problem we illustrate a pitfall which we will have to avoid. Note that

$$x^2 = x + x + \cdots + x \qquad (x \text{ terms}) \tag{3}$$

whenever x is a positive integer, since the right-hand is merely x times x. Differentiating we get $2x=1+1+\cdots+1$ (x terms) $=x$ so that, taking $x=1$, we get $2=1$.

The fallacy is that, since (3) holds only when x is a positive integer, it makes no sense to differentiate it. What is needed for the standard differentiation formulas is equality of the functions involved throughout an interval.

2. Computation of $P_n(x)$. Our object is now to prove:

THEOREM 2. *For each positive integer n there is a number C_n (depending on n, but not on x) such that*

$$P_n(x) = n\int_0^x P_{n-1}(t)dt + C_n x. \tag{4}$$

The proof will require:

LEMMA. *Let $f(x)$ and $g(x)$ be polynomials (in x) of degree $\leq d$ which are equal for more than d values of x. Then $f(x)$ and $g(x)$ are the* same polynomial. *In particular, they are equal for all values of x.*

Proof. Let $f(x)=a_dx^d+a_{d-1}x^{d-1}+\cdots+a_1x+a_0$ and $g(x)=b_dx^d+b_{d-1}x^{d-1}+\cdots+b_1x+b_0$. Then

$$h(x)=f(x)-g(x)=(a_d-b_d)x^d+(a_{d-1}-b_{d-1})x^{d-1}+\cdots+(a_1-b_1)x+(a_0-b_0)$$

is polynomial of degree $\leq d$ such that the *equation* $h(x)=0$ has more than d roots, name y all values of x for which $f(x)=g(x)$. But it is a well-known theorem that an equation of the form $c_dx^d+c_{d-1}x^{d-1}+\cdots+c_1x+c_0=0$ (each coefficient c_i a real number) has at most d real roots unless $0=c_d=c_{d-1}=\cdots=c_1=c_0$. (See, for example, E. Swokowski: *Fundamentals of College Algebra*, p. 231, Corollary 8.25.) We conclude that the coefficient a_i-b_i of each power x^i of x in $h(x)$ equals zero. In other words, each $a_i=b_i$, which is what we wanted to prove.

Proof of the theorem. By Theorem 1, $P_n(x)=1^n+2^n+\cdots+x^n$. Subtracting from this $P_n(x-1)=1^n+2^n+\cdots+(x-1)^n$ we get

$$(5) \qquad P_n(x) - P_n(x-1) = x^n$$

whenever x is a positive integer. Writing $n-1$ in place of n in (5) and multiplying both sides by n, we get $n[P_{n-1}(x) - P_{n-1}(x-1)] = nx^{n-1}$. Thus, differentiating both sides of (5), we see

$$(6) \qquad P_n'(x) - P_n'(x-1) = n[P_{n-1}(x) - P_{n-1}(x-1)].$$

But wait! Are we again on our way to proving $2=1$? Fortunately not this time: Consider n fixed temporarily. Then $P_n(x)$ is a polynomial of degree $n+1$ in x (Theorem 1). Thus both sides of (5) are polynomials in x of degree $\leqq n+1$ which are equal for more than $n+1$ values of x. (They are equal for all positive integers x.) The lemma then tells us that the left and right sides of (5) are the *same polynomial*, written in two different ways. Thus we are justified in differentiating (5) to obtain (6); and (6) will be true *for all real numbers* x (and all positive integers n).

Writing $x-1$ for x in (6) we get

$$P_n'(x-1) - P_n'(x-2) = n[P_{n-1}(x-1) - P_{n-1}(x-2)]$$

which, when added to (6) gives $P_n'(x) - P_n'(x-2) = n[P_{n-1}(x) - P_{n-1}(x-2)]$. Writing $x-2$ for x in (6) and adding the result to our last equation gives

$$P_n'(x) - P_n'(x-3) = n[P_{n-1}(x) - P_{n-1}(x-3)].$$

Proceeding in this way, we eventually obtain

$$(7) \quad P_n'(x) - P_n'(y) = n[P_{n-1}(x) - P_{n-1}(y)] \quad \text{whenever } x - y = \text{positive integer}.$$

Note that, in (7), y does not have to be positive, nor does it have to be an integer.

Now specialize (7) by taking x to be a positive integer. Then $x-0$ is a positive integer, so we can choose $y=0$ in (7), getting

$$P_n'(x) = nP_{n-1}(x) + [P_n'(0) - nP_{n-1}(0)].$$

Since we don't know the values of $P_n'(0)$ or $P_{n-1}(0)$, denote the expression in square brackets by C_n (to indicate that it depends on n but not on x). Then this last equation takes the form

$$8) \qquad P_n'(x) = nP_{n-1}(x) + C_n \quad (x \text{ and } n \text{ positive integers}).$$

Again consider n fixed. Since both sides of (8) are polynomials in x of degree $\leqq n$ which are equal for all positive integers x, the lemma tells us that (8) holds for all real numbers x. Therefore we can integrate both sides from 0 to x (after replacing x by a dummy variable t), getting

$$P_n(x) - P_n(0) = n\int_0^x P_{n-1}(t)dt + C_n x.$$

This is precisely what we want, provided we can show $P_n(0)=0$. But putting $x=1$ in (5) gives $1^n - P_n(0) = 1^n$; and this completes the proof of the theorem.

3. Factorization of $P_n(x)$. In order to find some of the factors of $P_n(x)$ we shall need some information about its values when x is not a positive integer. To obtain this we prove:

THEOREM 3. *For every positive integer n and every real number x,*

$$P_n(-x) = (-1)^{n+1}P_n(x-1). \tag{9}$$

REMARK. Before proving the theorem we observe that it can be interpreted geometrically: *If n is odd, the graph of $y=P_n(x)$ is symmetric with respect to reflection in the line $x=-\frac{1}{2}$; while if n is even it is symmetric with respect to reflection in the point $(-\frac{1}{2}, 0)$.* To see this, note that the number midway between $-x$ and $x-1$ is $\frac{1}{2}[-x+(x-1)]=-\frac{1}{2}$.

Proof of the Theorem. In equation (5), write $-x+1$ in place of x (which we can do since (5) is true for all real numbers x) and interchange the order of terms on the left-hand side:

$$-P_n(-x) + P_n(-x+1) = (-x+1)^n \quad \text{(for every real number } x\text{)}. \tag{10}$$

Note that $P_n(-x)$ cannot be interpreted as the sum of a series if $-x<0$. It is merely the value of a polynomial function.

PROBLEM 1. Finish the proof of Theorem 3. (Let x be a positive integer in (10); then successively write $-x+1$, $-x+2$, $-x+3, \cdots, -2, -1$ for $-x$ in (10) and add all the resulting equations.)

COROLLARY 1 (To Theorem 3). *$C_n=0$ whenever n is odd and >1.*

Proof. When n is odd, Theorem 3 becomes $P_n(-x)=P_n(x-1)$. Therefore Theorem 2 shows

$$P_n(x-1) = P_n(-x) = n\int_0^{-x} P_{n-1}(u)du - C_nx \tag{11}$$

make the change of variable $u=-t$ under the integral sign in (11).

$$P_n(x-1) = -n\int_0^{x} P_{n-1}(-t)dt - C_nx.$$

Now use Theorem 3 under the integral sign, with $n-1$ in place of n, noting that $(-1)^{(n-1)+1}=(-1)^n=-1$ since n is odd. (Here we use the fact that $n>1$; for if $n=1$, $n-1$ is not a positive integer, and hence Theorem 3 cannot be used.)

$$P_n(x-1) = n\int_0^{x} P_{n-1}(t-1)dt - C_nx. \tag{12}$$

Now subtract (12) from $P_n(x)=n\int_0^x P_{n-1}(t)dt+C_nx$, recalling that $P_n(x)-P_n(x-1)=x^n$ for all real numbers x (by (5)).

$$x^n = n\int_0^{x} t^{n-1}dt + 2C_nx = x^n + 2C_nx \quad (\textit{for all real } x).$$

Finally, putting $x=1$ we see that $C_n=0$, as desired.

COROLLARY 2. *Each polynomial* $P_n(x)$ *has the following factors*:

$$x(x+1) \qquad \text{for all } n. \tag{13}$$

$$x(x+1)(2x+1) \qquad \text{for all even } n. \tag{14}$$

$$x^2(x+1)^2 \qquad \text{for all odd } n>1. \tag{15}$$

Proof. To see that x is a factor of $P_n(x)$, it is sufficient to show that $P_n(0)=0$, which we obtained in the last sentence of Section 2.

To see that $x+1$ is a factor, set $x=0$ in Theorem 3, getting $0=\pm P_n(-1)$ and proving (13).

Next, set $x=1/2$ in Theorem 3: $P_n(-1/2)=(-1)^{n+1}P_n(-1/2)$. If n is *even* we get $2P_n(-1/2)=0$ so that $x+(1/2)$ is a factor of $P_n(x)$. Clearing fractions we see that $2x+1$ is a factor of $P_n(x)$. Thus we have (14).

To obtain (15) we need a lemma which seems to have disappeared from recent calculus courses:

LEMMA. *Let* $f(x)$ *be a polynomial and* c *a real number. Then* $(x-c)^2$ *is a factor of* $f(x)\Leftrightarrow f(c)=0$ *and* $f'(c)=0$.

Proof. ($\Leftarrow$). Since $f(c)=0$, $x-c$ is a factor of $f(x)$, say $f(x)=(x-c)g(x)$. Then $f'(x)=g(x)+(x-c)g'(x)$. Substituting $x=c$ we see that $0=g(c)$ so that $x-c$ is a factor of $g(x)$. Hence $(x-c)^2$ is a factor of $f(x)$.

($\Rightarrow$). Write $f(x)=(x-c)^2g(x)$ so that $f'(x)=2(x-c)g(x)+(x-c)^2g'(x)$. Letting $x=c$ in these equations completes the proof.

Now, differentiate Theorem 2: $P_n'(x)=nP_{n-1}(x)+C_n$. When n is odd, and >1, $C_n=0$ (corollary 1), so that $P_n'(x)=nP_{n-1}(x)$. We already know that $P_n(0)=P_n(-1)=0$ for every positive integer n. Hence $P_n'(0)=P_n'(-1)=0$ for every odd positive integer $n>1$. Thus, by the lemma, x^2 and $(x+1)^2$ are factors of $P_n(x)$.

CAUTION. Note that $1^4+2^4+3^4=98$ which is *not* divisible by $x(x+1)(2x+1)=3\times(3+1)\times(2\times 3+1)=84$, an apparent contradiction to (14) of Corollary 2. The reason for this apparent contradiction is that we must distinguish between arithmetical and algebraic divisibility. Thus, the assertion (14) of Corollary 2 that $P_n(x)$ is divisible by $x(x+1)(2x+1)$ when n is even means that for such an n there is a polynomial $g(x)$ such that $P_n(x)=x(x+1)(2x+1)g(x)$. In fact, from the table under Theorem 1,

$$P_4(x) = x(x+1)(2x+1)\left(\frac{x^2}{10}+\frac{x}{10}-\frac{1}{30}\right).$$

Because of the fractional coefficients in the last factor, we cannot conclude that the *number* $P_4(x)$ is divisible by $x(x+1)(2x+1)$. However, we *can* conclude that the number $30\,P_4(x)$ is divisible by the number $x(x+1)(2x+1)$.

PROBLEM 2. Find a positive integer m such that, for all positive integers x, $m(1^4+2^4+\cdots+x^4)$ is divisible by $1^2+2^2+\cdots+x^2$. (Better: find the smallest such m.) Generalize the solution of problem 2 to a proof of the following curious consequence of Corollary 2:

COROLLARY 3. *For every even positive integer n there is a positive integer M_n such that (for all x) $M_n(1^n+2^n+\cdots+x^n)$ is divisible by* $1^2+2^2+\cdots+x^2$.

In closing, we remark that in order to understand Corollary 3 one only has to know about the positive integers. Yet its *proof* involves properties of differentiable functions, hence requires a knowledge of the entire real number system.

4. Remarks and Problems. The integral formula for $P_n(x)$ (Theorem 2) makes it easy to calculate a table of these polynomials. However, the proof of this formula required the prior knowledge that the polynomials $P_n(x)$ exist for every positive integer n (Theorem 1). Thus the solution of the following problem is essential to complete the proof of Theorem 2.

PROBLEM 3. (a) Starting with $x^4-(x-1)^4=4x^3-6x^2+4x-1$, and writing successively $x-1$, $x-2$, $\cdots$, 3, 2, 1 in place of x, derive the formula

$$1^3+2^3+\cdots+x^3=\tfrac{1}{4}[x^4+6(1^2+2^2+\cdots+x^2)-4(1+2+\cdots+x)+x]$$

and use this, together with the formulas for $P_1(x)$ and $P_2(x)$ given in the table, to find $P_3(x)$.

(b) Generalize problem (3a) to a proof of Theorem 1 (mathematical induction will be necessary).

PROBLEM 4(a) Is the coefficient of x^{n+1} in $P_n(x)$ always $1/(n+1)$?

(b) Is the coefficient of x^n in $P_n(x)$ always $1/2$?

(c) Is the coefficient of x^{n-1} in $P_n(x)$ always $n/12$?

(d) Is it true of $P_n(x)$ that, except for the term of degree n, all terms have even degree, or else all have odd degree?

PROBLEM 5. Show that $P_n'(17)-nP_{n-1}(17)=P_n'(23)-nP_{n-1}(23)$ for every positive integer n.

PROBLEM 6. Show that, if $a_x=c_nx^n+c_{n-1}x^{n-1}+\cdots+c_1x+c_0$, the c_i's being constants, then

$$a_1+a_2+\cdots+a_x=c_nP_n(x)+c_{n-1}P_{n-1}(x)+\cdots+c_1P_1(x)+c_0x.$$

Use this fact, and our table of $P_n(x)$ to find the sum of

$$1^2+3^2+5^2+\cdots+(2x-1)^2$$
$$(1\cdot2\cdot3)+(2\cdot3\cdot4)+(3\cdot4\cdot5)+\cdots+x(x+1)(x+2).$$

The asterisks below indicate more challenging problems.

PROBLEM 7* (by William Self). Compute $\int_{-1}^{0}P_n(x)P_m(x)dx$ when m is even and n is odd ($m, n\geqq 2$). (Hint: the answer will be the same for all such m and n.)

PROBLEM 8.* Note that $P_4(x)$ and $P_5(x)$ can be factored completely into linear factors with real number coefficients. Does $P_6(x)$ have any real linear

factors other than those listed in the table of $P_n(x)$?

PROBLEM 9.* For each n, the graph of the function $y=P_n(x)$ "interpolates" the sequence of points $(1, 1^n)$, $(2, 1^n+2^n)$, $(3, 1^n+2^n+3^n)$, . . . by connecting them with a continuous curve. Is this interpolation genuine in the sense that, when $x \geqq 1$, the curve $y=P_n(x)$ is always concave upward? (Hint: First observe that when n is even $P_n''(x)=n(n-1)P_{n-2}(x)$; and when n is odd $P_n'(x)=nP_{n-1}(x)$.)

We have not obtained much information about the numbers C_n (n even). For example: is C_n always nonzero when n is even, do the numbers C_2, C_4, $C_6, \cdots$ alternate in sign, do they approach a limit as $n \to \infty$? The answers to these questions are not all what one would expect. Some additional computations would show:

$$C_2 = \frac{1}{6} \qquad C_4 = -\frac{1}{30} \qquad C_6 = \frac{1}{42} \qquad C_8 = \frac{-1}{30} \qquad C_{10} = \frac{5}{66}$$

$$C_{12} = \frac{-691}{2730} \qquad C_{14} = \frac{7}{6} \qquad C_{16} = \frac{-3617}{510} \qquad C_{18} = \frac{43867}{798} \qquad C_{20} = \frac{-122277}{2310}.$$

(Source: G. Boole, *Treatise on the Calculus of Finite Differences*, p. 90 of the third edition (S. T. Stechert and Co., N. Y., 1872). In Boole's notation, our $C_n = (-1)^{(n+2)/2}B_{n-1}$ (n even) where B_{n-1} is the $(n-1)$'st "Bernoulli" number.) Boole shows, in fact, that

$$\lim_{n \text{ even},\, n\to\infty} \frac{(-1)^{(n+2)/2}C_n}{(2(n)!/(2\pi)^n)} = 1$$

so that, as $n \to \infty$ (n even), C_n/n^2C_{n-2} approaches $-1/4\pi^2$ (same source, chap. 6, p. 109). This answers all the above questions and suggests:

PROBLEM 10.** Can any of the above properties of C_n be obtained by our methods? (The double asterisk indicates that I don't know the answer.)

PROBLEM 11.** $h(x)=3x^4+6x^3-3x+1$ is a factor of $P_6(x)$ (see the table of $P_n(x)$). One factor of $h'(x)$ is $2x+1$, and $h'(x)/(2x+1)=3(2x^2+2x-1)=$ a factor of $P_5(x)$. Is this a special case of some general property of $P_n(x)$?

PROBLEM 12.** When $n \geqq 6$, does $P_n(x)$ have any linear factors other than those listed in Corollary 2 to Theorem 3? (Note: It is clear that $P_4(x)$ and $P_5(x)$ have additional linear factors; and problem 8 is the special case of this question for $n=6$. Part of the answer to problem 9 will be that $P_n(x)=0$ has no real roots $\geqq 1$; hence by the symmetry of $P_n(x)$ (Theorem 3), it has no real roots $\leqq -2$. Thus any linear factors must come from roots between -2 and $+1$.)

The author is indebted to Mr. Edward Shapiro of Lincoln High School, Brooklyn, N. Y., for his help in the preparation of this paper.

POWERFUL NUMBERS

S. W. GOLOMB, University of Southern California

1. Introduction. We define a positive integer r to be a *powerful number* if p^2 divides r whenever the prime p divides r. If $k(x)$ denotes the number of powerful numbers $\leqq x$, we show that $k(x) \sim c\sqrt{x}$, with the constant $c = \zeta(3/2)/\zeta(3)$, as $x \to \infty$. We also prove that there are infinitely many pairs of consecutive powerful integers, such as 8; 9 and 288; 289. We conclude with some results and conjectures concerning the gaps between powerful numbers.

2. The density of powerful numbers. Let

$$(1) \qquad F(s) = \prod_p (1 + p^{-2s} + p^{-3s} + p^{-4s} + \cdots) = \prod_p \left(1 + \frac{1}{p^s(p^s - 1)}\right),$$

where the products are extended over all primes p. It is evident that

$$(2) \qquad F(s) = \sum_{r \in K} r^{-s},$$

where K is the set of powerful numbers. Then, the sum of the reciprocals of the powerful numbers,

$$(3) \qquad F(1) = \sum_{r \in K} \frac{1}{r} = \prod_p \left(1 + \frac{1}{p(p-1)}\right),$$

is seen to be convergent.

To estimate $k(x)$, the number of powerful numbers up to x, we observe first that $k(x) \geqq [\sqrt{x}]$, since every perfect square is powerful. Next, we observe that every powerful number r can be represented as a perfect square n^2 (including the case $n = 1$) times a perfect cube m^3 (including $m = 1$), and that this representation is unique if we require m to be square-free. That is, we set m equal to the product of those primes having odd exponents in the canonical factorization of r into powers of distinct primes, and the representation $r = n^2m^3$ is then unique.

Thus,

$$(4) \quad k(x) = \#(n^2m^3 \leqq x,\, \mu(m) \neq 0) = \sum_{m=1}^{\infty} \mu^2(m) \left[\left(\frac{x}{m^3}\right)^{1/2}\right] \sim cx^{1/2}, \qquad x \to \infty,$$

where

$$(5) \qquad \sum_{m=1}^{\infty} \mu^2(m) m^{-3/2} < \zeta(3/2) < \infty.$$

Explicitly,

$$(6) \qquad c = \prod_p (1 + p^{-3/2}) = \prod_p (1 - p^{-3})/(1 - p^{-3/2}) = \zeta(3/2)/\zeta(3),$$

where $\zeta(s)$ is the Riemann Zeta Function (cf. [1], p. 5). This evaluation of c comes from setting $s=3/2$ in the identity

$$(7)\quad \sum_{m=1}^{\infty}\frac{\mu^2(m)}{m^s}=\prod_p\left(1+\frac{1}{p^s}\right)=\prod_p\frac{1-p^{-2s}}{1-p^{-s}}=\prod_p(1-p^{-2s})\Big/\prod_p(1-p^{-s})$$
$$=\zeta(s)/\zeta(2s)$$

for all $Rl(s)>1$, where $\zeta(s)=\sum_{n=1}^{\infty}n^{-s}=\Pi_p(1-p^{-s})^{-1}$ for $Rl(s)>1$.

For purposes of estimation, we have the inequalities

$$(8)\qquad cx^{1/2}\geqq k(x)\geqq cx^{1/2}-3x^{1/3}\qquad \text{for } x\geqq 1,$$

because $cx^{1/2}=\sum_{m=1}^{\infty}\mu^2(m)(x/m^3)^{1/2}\geqq\sum_{m=1}^{\infty}\mu^2(m)\ [(x/m^3)^{1/2}]=k(x)$, and

$$cx^{1/2}-k(x)=\sum_{m=1}^{\infty}|\mu(m)|\left\{\left(\frac{x}{m^3}\right)^{1/2}-\left[\left(\frac{x}{m^3}\right)^{1/2}\right]\right\}\leqq\sum_{m=1}^{[x^{1/3}]-1}|\mu(m)|\cdot 1$$
$$+\sum_{m=[x^{1/3}]}^{\infty}|\mu(m)|\left(\frac{x}{m^3}\right)^{1/2}\leqq([x^{1/3}]-1)+\left(1+\sqrt{x}\int_{[x^{1/3}]}^{\infty}u^{-3/2}du\right)$$
$$\leqq x^{1/3}+2x^{1/2}\cdot x^{-1/6}=3x^{1/3}.$$

Numerically, $c=2\cdot 173\cdots$.

We have the further identities:

$$(9)\qquad \begin{aligned} F(s)&=\sum_{r\in K}(1/r^s)=\sum_{n=1}^{\infty}n^{-2s}\cdot\sum_{m=1}^{\infty}\mu^2(m)m^{-3s}\\ &=\sum_{t=1}^{\infty}t^{-2s}\sum_{m\mid t}|\mu(m)|/m^s=\sum_{t=1}^{\infty}t^{-2s}\prod_{p\mid t}(1+p^{-s}),\end{aligned}$$

where we used the substitution $t=mn$;

$$(10)\qquad F(s)=\sum_{n=1}^{\infty}n^{-2s}\cdot\sum_{m=1}^{\infty}\mu^2(m)m^{-3s}=\zeta(2s)\zeta(3s)/\zeta(6s),$$

and

$$(11)\qquad F(1)=\zeta(2)\zeta(3)/\zeta(6)=\sum_{t=1}^{\infty}\Psi(t)/t^3,$$

where

$$(12)\qquad \Psi(t)=t\prod_{p\mid t}\left(1+\frac{1}{p}\right),$$

by setting $s=1$ in the previous identities (9) and (10).

Since $\zeta(2)=\pi^2/6$ and $\zeta(6)=\pi^6/945$, we observe

$$(13)\qquad F(1)=\frac{315}{2\pi^4}\zeta(3).$$

3. Consecutive powerful numbers. Four consecutive integers cannot all be powerful, since one of them is twice an odd number. No example of three consecutive powerful numbers is known, unless one is willing to accept -1, 0, 1. If such an example exists, it must be of the form

$$4k-1, \qquad 4k, \qquad 4k+1.$$

No case of $4k-1$ and $4k+1$ both being powerful is known. In fact, the only known example of consecutive odd numbers $2k-1$ and $2k+1$ both being powerful is $2k-1=25$, $2k+1=27$.

There are two infinite families of examples where two consecutive integers are powerful which correspond to the solutions of the Pell equations $x^2-2y^2=1$ and $x^2-2y^2=-1$. (Each of these equations is well known to have infinitely many solutions [2].)

Let x_1, y_1 satisfy $x_1^2-2y_1^2=\pm 1$. Then $8x_1^2y_1^2=A$ and $(x_1^2+2y_1^2)^2=B$ are consecutive powerful numbers! In Table I, we see several examples of this type.

TABLE I. Consecutive powerful numbers from solutions of the Pell equations $x^2-2y^2=\pm 1$.

x	y	A	B
1	1	$8=2^3$	$9=3^2$
3	2	$288=2^5\cdot 3^2$	$289=17^2$
7	5	$9800=2^3\cdot 5^2\cdot 7^2$	$9801=3^4\cdot 11^2$
17	12	$332{,}928=2^5\cdot 3^2\cdot 17^2$	$332{,}929=577^2$
$\sqrt{B_0}$	$\sqrt{A_0/2}$	$4A_0B_0$	$4A_0B_0+1$

If A and $B=A+1$ are consecutive powerful numbers, and if B is a perfect square, $B=u^2$, then $A=(u-1)(u+1)$. If u is even, then $(u-1, u+1)=1$, and both $u-1$ and $u+1$ are odd powerful numbers. As already remarked, the only known instance of this phenomenon is $u-1=25$, $u+1=27$, leading to the isolated example $A=675=3^3\cdot 5^2$, $B=676=2^2\cdot 13^2$. If u is odd, then $(u-1)/2$ and $(u+1)/2$ are consecutive integers, with $((u-1)/2, (u+1)/2)=1$. For u^2-1 to be powerful, $(u-1)/2$ and $(u+1)/2$ must be a powerful odd number and twice a powerful number, in either order. The two Pell equations produce examples in both orders. However, an example satisfying neither of these Pell equations is also known, with $(u-1)/2=242=2\cdot 11^2$ and $(u+1)/2=243=3^5$. This leads to $A=235{,}224=2^3\cdot 3^5\cdot 11^2$ and $B=235{,}225=5^2\cdot 97^2$.

Whenever A and B are consecutive powerful numbers, so too are $A'=4AB$ and $B'=4AB+1=(2A+1)^2$. The solution $x_0=1$, $y_0=1$, of $x^2-2y^2=-1$ generates all solutions of the Pell equations $x^2-2y^2=\pm 1$, in the sense that $x_n+y_n\sqrt{2}=(x_0+y_0\sqrt{2})^n$ yields the complete set of solutions (x_n, y_n) such that $x_n^2-2y_n^2=\pm 1$.

Note that the consecutive powerful numbers $A=675$, $B=676$, come from the solution $x=26$, $y=15$, of the Pell equation $x^2-3y^2=1$, with $A=3y^2$ and

$B=x^2$. Similarly, the example $A=235{,}224$, $B=235{,}225$ of consecutive powerful numbers arises from the Pell equation $x^2-6y^2=1$ with $x=485$, $y=198$. More generally, any solution x_1, y_1 of the Pell equation $x^2-dy^2=\pm 1$, with the extra condition that $d\mid y_1^2$, leads to an infinite family of consecutive powerful numbers, starting with $A_1=x_1^2$, $B_1=dy_1^2=A_1\pm 1$, and continuing with $A_n=x_n^2$, $B_n=dy_n^2$, where (x_n, y_n) are obtained from the computation $(x_1+\sqrt{d}y_1)^n=x_n+\sqrt{d}y_n$.

Conversely, whenever we have two consecutive powerful integers, if one of them is a perfect square x^2, we can write the other in the form $n^2m^3=my^2$, with m square-free, and we have a solution to the Pell equation $x^2-my^2=\pm 1$.

In all cases given thus far of consecutive powerful numbers, the *larger* number is a perfect square. However, the Pell equation $x^2-5y^2=-1$ with $5\mid y$ leads to infinitely many powerful numbers $x^2+1=5y^2$, such as $x^2=(682)^2=465{,}124$; $5y^2=5(305)^2=5^3\cdot 61^2=465{,}125$.

One example has been found consisting of two consecutive powerful numbers where *neither* is a perfect square: $A=23^3=12{,}167$ and $B=2^3\cdot 3^2\cdot 13^2=12{,}168$. It is not known how many such examples (without perfect squares) exist altogether.

4. Gaps between powerful numbers. The set K of powerful numbers is closed under multiplication. Since there are infinitely many pairs of powerful numbers which differ by 1, there are infinitely many pairs of powerful numbers differing by r, for any $r\in K$.

Every positive integer not of the form $2(2b+1)$ is a difference of two powerful numbers in at least one way (specifically, as a difference of two perfect squares). For numbers of the form $2(2b+1)$, $b\geqq 0$, such representations may also exist. Thus:

$2=3^3-5^2$	$30=83^2-19^3$
$6=?$	$34=?$
$10=13^3-3^7$	$38=37^2-11^3$
$14=?$	$42=?$
$18=19^2-7^3=3^2(3^3-5^2)$	$46=17^2-3^5$
$22=7^2-3^3=47^2-3^7$	$50=5^2(3^3-5^2)$
$26=3^3-1^k=7^2\cdot 3^5-109^2$	$54=3^4-3^3=3^3(3^3-5^2)$
	$=7^3-17^2$.

If u and v are both powerful numbers, $(u, v)=1$, and $a=u-v$, we say that a has a *proper* representation as a difference of powerful numbers. We observe:

$$2b+1=(b+1)^2-b^2$$
$$8c=(2c+1)^2-(2c-1)^2$$

so that all odd numbers, as well as all multiples of 8, have proper representa-

tions. Among the numbers $2(2b+1)$ on the range $2(4)54$ for which representations were found, there were proper representations included in every case except $2(2b+1)=50$. Finally, for numbers $4(2b+1)$ on the range $4(8)100$, we observe the following proper representations:

$4 = 5^3 - 11^2$	$60 = ?$
$12 = 47^2 - 13^3$	$68 = 3^3 \cdot 5^4 - 7^5$
$20 = ?$	$76 = 5^3 - 7^2$
$28 = ?$	$84 = ?$
$36 = ?$	$92 = ?$
$44 = 5^3 - 3^4 = 13^2 - 5^3$	$100 = 7^3 - 3^5.$
$52 = ?$	

It is interesting that if u and $v=u+4$ are both powerful, then so too are $u'=uv$ and $v'=u'+4=(u+2)^2$. Thus, from the example $4=5^3-11^2$, an infinite number of proper representations of 4 are obtainable. It would be interesting to determine whether or not any numbers other than 1 and 4 have infinitely many proper representations.

Among the powerful numbers which are not perfect squares, the smallest difference known to occur infinitely often is 4. Specifically, the equation $3x^2 - 2y^2=1$ has infinitely many solutions for which $3|x$, such as $x=9$, $y=11$. For any such solution, we have $12x^2-8y^2=4$, where $12x^2$ and $8y^2$ are both powerful, and neither is a square. The only known instances where the difference between nonsquare powerful numbers is less than 4 are: $2^7-5^3=3$, and (as previously mentioned) $2^3 \cdot 3^2 \cdot 13^2-23^3=1$.

It has been conjectured that 6 cannot be represented in any way as a difference between two powerful numbers. It is further conjectured that there are infinitely many numbers which cannot be so represented.

This work was supported in part by the Air Force Office of Scientific Research under Grant AFOSR-68-1555.

References

1. E. C. Titchmarsh, The Theory of the Riemann Zeta Function, Clarendon Press, Oxford, England, 1951.

2. W. J. Le Veque, Topics in Number Theory, Vol. I, Ch. 8, Addison-Wesley, Reading, Mass., 1956, pp. 139–144.

EDITED BY DAVID DRASIN

Manuscripts for this Department should be sent to David Drasin, Division of Mathematica Sciences, Purdue University, Lafayette, IN 47907.

POLYNOMIALS WITH PREASSIGNED VALUES AT THEIR BRANCHING POINTS

JAN MYCIELSKI, University of Colorado

It is the purpose of this paper to give a simple proof of the following

THEOREM. *Given a sequence of n complex numbers $a_1, \cdots, a_n$ there exists a polynomial $P(x)$ of degree $n+1$ and n complex numbers $b_1, \cdots, b_n$ such that the highest coefficient of P is 1, $P(0)=0$, $P(b_i)=a_i$ and $P'(b_i)=0$ for $i=1, \cdots, n$. Moreover* (*) *if β occurs k times in the sequence $b_1, \cdots, b_n$ then $(x-\beta)^{k+1}$ divides $P(x)-P(\beta)$.*

This theorem was announced by J. W. Andrushkiw [0] (in a different formulation) and proved by R. Thom [8]. Before [8] proofs were published for the case when a_i are real and satisfy some additional conditions, see A. J. Kempner [4], Chandler Davis [1], J. Mycielski and S. Paszkowski [6] and W. J. Kammerer [3]. See also V. S. Videnskyj [10], [11], Chandler Davis [2] and R. Thom [7] for related studies.

In [1], [6] and [3] it was shown that if a_i are real and satisfy some additional conditions then P is unique. In our case P is not unique but we can see analysing the proof which follows that if the numbers $0, a_1, a_2, \cdots, a_n$ are all different then the number N of such polynomials is a finite constant depending only on n. I conjecture that $N=(n+1)^n$.

By the theorem of Tarski on the elementary equivalence of algebraic closed fields of a given characteristic (see e.g. [5] or [9]) the theorem extends to all algebraic closed fields of characteristic 0. Our proof uses the topology of the field of complex numbers. Perhaps if a purely algebraic proof was found it would yield some extensions to other fields e.g. of characteristic $\neq 0$.*

Proof of the Theorem. For any point $b=(b_1, \cdots, b_n)$ of the n-dimensional complex space $\boldsymbol{C}^n$ we put

$$(1) \qquad P_b(x) = (n+1)\int_0^x \prod_{i=1}^n (t-b_i)dt$$

and we define the vector-valued function

$$(2) \qquad \varphi(b) = (P_b(b_1), \cdots, P_b(b_n)).$$

* *Note added on December 8, 1969.* Recently R. E. MacRae found such a proof. Applying the theorem of Bézout he generalizes the theorem proved in this paper and his generalization is valid for algebraic closed fields of characteristic p if p does not divide $n+1$ and implies my conjecture that $N=(n+1)^n$.

Of course it is enough to prove that φ maps C^n *onto* itself and that (*) holds.

First let us check (*). Indeed it follows directly from (1) that

$$d^s(P_b(x) - P_b(\beta))/dx^s$$

is divisible by $x-\beta$ for $s=0, \cdots, k$, i.e., (*) is true.

It follows from (*) that

$$\text{(3)} \qquad \text{if} \quad \varphi(b) = (0, \cdots, 0) \quad \text{then} \quad b_1 = b_2 = \cdots = b_n = 0.$$

Indeed (*) implies then that $P(x)=x(x-b_1)\cdots(x-b_n)$ but this is inconsistent with (1) unless $b_1=b_2=\cdots=b_n=0$.

It is clear from (2) that φ is homogeneous of degree $n+1$, i.e.,

$$\text{(4)} \qquad \varphi(\alpha b) = \alpha^{n+1}\varphi(b) \qquad \text{for } \alpha \in C.$$

Now we will prove that

(5) φ is a local homeomorphism at each point b such that $0, b_1, \cdots, b_n$ are all different.

This is done similarly as in [1] but somewhat simpler. Namely we will show that the Jacobian determinant of φ does not vanish at such points. Of course the Jacobian matrix $(\partial P_b(b_r)/\partial b_s)$ is given by

$$\text{(6)} \qquad \frac{\partial P_b(b_r)}{\partial b_s} = -(n+1)\int_0^{b_r} \prod_{i=1, i\neq s}^{n} (t-b_i)dt \qquad (r, s = 1, \cdots, n).$$

Suppose to the contrary that $\det(\partial P_b(b_r)/\partial b_s)=0$, i.e., that there are constants $c_1, \cdots, c_n$ not all 0 such that

$$\text{(7)} \qquad \sum_{s=1}^{n} c_s \frac{\partial P_b(b_r)}{\partial b_s} = 0 \qquad \text{for } r = 1, \cdots, n.$$

Clearly,

$$\sum_{s=1}^{n} c_s \prod_{i=1, i\neq s}^{n} (t-b_i)$$

is a polynomial which does not vanish for $t=b_i$ if $c_i\neq 0$ and hence the function

$$F(x) = -(n+1)\int_0^{x} \sum_{s=1}^{n} c_s \prod_{i=1, i\neq s}^{n} (t-b_i)dt$$

is a polynomial of positive degree $\leqq n$. But (6) and (7) yield that F vanishes at $n+1$ points $0, b_1, \cdots, b_n$ which is a contradiction. Thus (5) is proved.

Now let us replace $(b_1, \cdots, b_n)$ by a variable $[b_0, \cdots, b_n]$ running over the complex n-dimensional projective space P_C^n (i.e., $b_0, \cdots, b_n$ are complex numbers not all 0 and $[b_0, \cdots, b_n]$ denotes the set of all sequences $(cb_0, cb_1, \cdots, cb_n)$ with $c\neq 0$). And modify φ to a map $\varphi^*: P_C^n \to P_C^n$ putting

$$\varphi^*[b_0, \cdots, b_n] = [b_0^{n+1}, P_b(b_1), \cdots, P_b(b_n)].$$

Clearly by (3) and (4) this is a correct definition of a continuous mapping φ^* of $P^n_{\mathbf{C}}$ into itself and φ^* becomes an extension of φ if we identify $(b_1, \cdots, b_n)$ with $[1, b_1, \cdots, b_n]$.

The further argument rests on the following topological Lemma.

(L) *Let f be a continuous mapping of a compact Hausdorff space X into a space Y and $B \subseteq X$ be a nonempty set such that $f(B)$ is open and the set $Y - f(X-B)$ is connected and dense in Y. Then f maps X onto Y.*

Proof. X being compact its image is closed in Y, hence it will be enough to check that $f(X)$ is dense in Y. $f(B)$ is open and the boundary of $f(B)$ is contained in $f(X-B)$. Therefore the boundary of $f(B)$ does not disconnect Y and hence $f(B)$ is dense in Y.

Now we apply (L) putting $f=\varphi^*$, $X=Y=P^n_{\mathbf{C}}$, $B=\{[1, b_1, \cdots, b_n]: 0, b_1, \cdots, b_n \text{ are all different}\}$. By (5) and since $\varphi^*(P^n_{\mathbf{C}}-B) \subseteq P^n_{\mathbf{C}}-B$ and B is open connected and dense in $P^n_{\mathbf{C}}$ all the suppositions of (L) are satisfied. Hence φ^* maps $P^n_{\mathbf{C}}$ onto itself. Let $A=\{[1, b_1, \cdots, b_n]: (b_1, \cdots, b_n) \in \mathbf{C}^n\}$, then $\varphi^*(P^n_{\mathbf{C}}-A) \subseteq P^n_{\mathbf{C}}-A$. Hence $\varphi^*(A)=A$ and $\varphi(\mathbf{C}^n)=\mathbf{C}^n$.

REMARK. Some a_i in the theorem may be equal but the corresponding b_i need not be equal. Indeed formula (2) shows that $\varphi(1, 2, 3)=(-9, -8, -9)$.

References

0. J. W. Andrushkiw, Polynomials with prescribed values at critical points, Bull. Amer. Math. Soc., 62 (1956) 243.

1. Chandler Davis, Extrema of a polynomial, this MONTHLY, 64 (1957), 679–680, and A. W. Goodman, Editorial note, *ibid.* p. 680.

2. ———, Properties of maps of some Chebyshev systems, Dokl. Acad. Nauk, SSSR, 175 (1967) 280–283.

3. W. J. Kammerer, Polynomial approximations to finitely oscillating functions, Math. of Computation, 15 (1961) 115–119.

4. A. J. Kempner, On the shape of polynomial curves, Tohoku Math. Journal, Part I, 37 (1933) 347–352; Part II, 42 (1936) 318–330.

5. G. Kreisel and J. L. Krivine, Elements of mathematical logic, Amsterdam, 1967.

6. Jan Mycielski and S. Paszkowski, A generalization of Chebyshev polynomials, Bull. Acad. Polon. Sci. Série Math., Astr. et Phys., 8 (1960) 433–438.

7. R. Thom, La stabilité topologique des applications polynomials, Enseignement Math., 8 (1962) 24–33.

8. ———, L'équivalence d'une fonction différentiable et d'un polynome, Topology 3, Suppl. 2 (1965) 297–307.

9. R. L. Vaught, Applications of the Löwenheim-Skolem-Tarski theorem to problems of completeness and decidability, Nederl. Akad. Wetensch. Proc. Ser. A, 57 (Indag. Math. vol. 16) (1954) 467–472.

10. V. S. Videnskyj, O postroyenii mnohochlena po dannoi posledovatelnosti yeho ekstremov, Mezhdunarodnaya Konferencija po teorii anal. funkcii, Yerevan 1965, p. 32, Moscow 1966.

11. ———, Theorem about the existence of a polynomial having a given sequence of extrema, Dokl. Acad. Nauk SSSR, 171 (1966) 17–19, (transl. Soviet Math. Dokl. (1966) 1395–1398.)

MORE BIRTHDAY SURPRISES

MORTON ABRAMSON, York University, Toronto, and W. O. J. MOSER, McGill University, Montreal

There does not seem to appear in the literature a solution to the following generalization of the well-known birthday problem. Suppose a "year" has n days. If $k \geqq 1$ is given, and p people are chosen at random, what is the probability $B_k(n, p)$ that the birthdays of every pair are at least k days apart? We obtain the answer as a special case of a more general result. For other birthday problems see [1], [2]. Note that the first and last days of the year are considered to be consecutive, so that, for example, with $n=365$, December 29 and January 3 are 5 days apart. Of course

$$B_1(365, p) = (365)(364) \cdots (365 - p + 1)/(365)^p$$

is the probability that (in an ordinary year) no two of the people have the same birthday.

The birthdays of the p people are described by a p-permutation

$$(1) \qquad b_1, b_2, \cdots, b_p \qquad b_i \in \{1, 2, \cdots, n\}, \quad b_i \neq b_j \text{ if } i \neq j.$$

We seek the number of permutations (1) satisfying

$$(2) \qquad k \leqq |b_i - b_j| \leqq r, \qquad 1 \leqq i \neq j \leqq p,$$

and this is $p!$ times the number of p-choices

$$(3) \qquad 1 \leqq x_1 < x_2 < \cdots < x_p \leqq n$$

satisfying

$$(4) \qquad k \leqq x_{i+1} - x_i, \qquad i = 1, 2, \cdots, p-1, \; x_p - x_1 \leqq r.$$

Corresponding to each p-choice (3) satisfying (4) there is a p-choice (obtained by setting $y_i = x_i - (i-1)(k-1)$, $i = 1, 2, \cdots, p$)

$$(5) \qquad 1 \leqq y_1 < y_2 < \cdots < y_p \leqq n - (p-1)(k-1)$$

satisfying

$$(6) \qquad y_p \leqq r + y_1 - (p-1)(k-1)$$

(obtained from $x_p - x_1 \leqq r$ upon replacing x_p by $y_p + (p-1)(k-1)$ and x_1 by y_1). We count these in the two cases: (I) $y_1 \geqq n-r$; (II) $y_1 < n-r$. If (I) holds, (6) is redundant and the p-choices are

$$n - r \leqq y_1 < y_2 < \cdots < y_p \leqq n - (p-1)(k-1);$$

there are

$$(7) \qquad \binom{n - (p-1)(k-1) - n + r + 1}{p}$$

of these. If (II) holds, the p-choices satisfy

$$1 \leqq y_1 < n - r, \qquad y_1 < y_2 < \cdots < y_p \leqq r + y_1 - (p-1)(k-1);$$

there are

$$(8) \qquad (n - r - 1)\binom{r + y_1 - (p-1)(k-1) - y_1}{p-1}$$

of these. Adding (7) and (8) we have the

$$(9) \qquad \frac{n - (p-1)(k-n+r)}{p}\binom{r - (k-1)(p-1)}{p-1}$$

p-choices (3) satisfying (4). (This proof has been included to make this note self-contained; the result is contained in [**3**; formula (37)′] and [**4**; formula (19) with l' large and $l = n - r$].) Hence,

$$(10) \qquad \frac{p!}{n^p}\cdot\frac{n - (p-1)(k-n+r)}{p}\binom{r - (k-1)(p-1)}{p-1}$$

is the probability that the birthdays (1) satisfy (2). With $r = n - k$ in (10) we have

$$B_k(n, p) = \frac{(p-1)!}{n^{p-1}}\binom{n - p(k-1) - 1}{p-1}.$$

The values of $1 - B_k(365, p)$ are listed for some values of p and k. For example, $1 - B_2(365, p)$ is the probability that (in an ordinary year) some pair have the same or adjacent birthdays. Letting $s(k)$ denote the smallest p for which $1 - B_k(365, p) \geqq \frac{1}{2}$ we see the surprisingly small number of people for which, with probability $\geqq \frac{1}{2}$, some pair have birthdays less than k days apart.

$1 - B_k(365, p)$

p \ k	1	2	3	4
5	.02713	.07971	.13013	.17844
8	.07433	.20873	.32604	.42812
9	.09462	.26042	.39901	.51433
10	.11694	.31472	.47209	.59648
11	.14114	.37056	.54328	.67210
12	.16702	.42693	.61090	.73952
13	.19440	.48287	.67363	.79778
14	.22310	.53749	.73053	.84663
21	.44368	.83603	.95537	.98890
22	.47569	.86378	.96774	.99313
23	.50729	.88791	.97709	.99586
24	.53834	.90864	.98401	.99757

k	$s(k)$
1	23
2	14
3	11
4	9
5	8
6	8
7	7
8	7
9	6
10	6

References

1. Robert E. Greenwood, and Arthur Richert, Jr., A birthday holiday problem, J. Combinatorial Theory, 5 (1968) 422–424.
2. M. S. Klamkin and D. J. Newman, Extensions of the birthday surprise, J. Combinatorial Theory, 3 (1967) 279–282.
3. R. Lagrange, Sur les combinaisons d'objets numérotés, Bull. Sci. Math., 87 (1963) 29–42.
4. W. O. J. Moser and Morton Abramson, Enumeration of combinations with restricted differences and cospan, J. Combinatorial Theory, 7 (1969) 162–170.

ON A PROBLEM OF J. C. MOORE

ALEX HELLER, City University of New York

J. C. Moore has proposed an interesting problem concerning the "contractibility" of continuous maps.

If X and Y are topological spaces, a map $f:X\rightarrow Y$ is *contractible* if X and Y have contracting homotopies $\xi_t:X\rightarrow X$ and $\eta_t: Y\rightarrow Y$ (i.e., homotopies of the identity with a constant map) such that $f\xi_t=\eta_t f$ for all t. Thus a contractible map has contractible domain and range. Moore asks whether the converse is true.

The following picture may suggest what is involved:

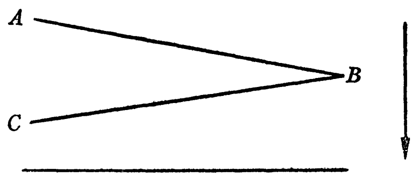


FIG. 1

The map is the projection indicated by the arrow. It is contractible: AB and CB need only be shrunk simultaneously to B.

Moore further suggests that the map $f:X\rightarrow Y$ indicated by the following picture:

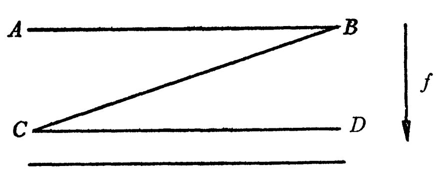


FIG. 2

is *not* contractible.

This suggestion seems plausible: it is hard to see how the homotopy ξ_t can deform the image of X away from A and D while B and C are covered. But

while A and D are covered so also, by the intermediate value theorem, are B and C. However, it seems difficult—and is for all I know impossible—to construct a proof along these lines.

The result may, however, be obtained by the following argument. If $g: U \to V$ define

$$\bar{U} = \{(u, u')/gu = gu'\} \subset U \times U \quad \text{and} \quad \bar{g}(u, u') = gu = gu'.$$

If the contracting homotopies $\theta_t: U \to U$ and $\sigma_t: V \to V$ give a contraction of g then $\bar{\theta}_t(u, u') = (\theta_t u, \theta_t u')$ defines a contracting homotopy of $\bar{U}$, and $\bar{g}\,\bar{\theta}_t = \sigma_t \bar{g}$, so that $\bar{g}$ is again contractible. In particular, $\bar{U}$ is contractible.

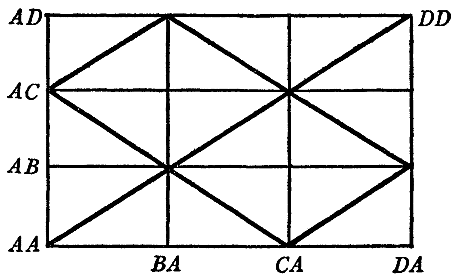


FIG. 3

Now let us apply this construction to the map f of Figure 2. Since X is an interval, $\bar{X}$ is a subspace of the square $X \times X$. One sees by inspection that it is the subspace indicated by the heavy lines in figure 3. But this is clearly not contractible, hence f is not contractible.

THE DYNAMICS OF AVALANCHES

D. R. ENGLISH, Rhodes University, Grahamstown, South Africa

(1) To simplify the problem, we assume that the mountainside where an avalanche takes place is an inclined plane of slope α to the horizontal, and that the snow is b thick. We also assume that snow is incompressible, with constant density ρ.

(2) An avalanche can happen in four different ways:

(i) The snow slides down the slope as a rigid body.

(ii) A hard lump of snow, ice, or rock rolls down the slope without gathering any snow.

(iii) The snow slides down the slope, collecting itself as it moves. The action is similar to that of a carpet pushed along one edge, crumpling without rolling.

(iv) The snow rolls down the slope, starting from the top. The action is similar to that of a carpet, wound up from one edge, without sliding.

(3) The actual behaviour of avalanches, as observed by the writer in India,

is a varying combination of the above motions. Also, snow is compressible, slopes vary, and air resistance acts. Therefore, an exact mathematical analysis of avalanche motion is not feasible. It is of interest, however, to calculate the acceleration in the above four cases, with the simplifying assumptions of paragraph (1). We shall refer to them in short as:

(i) Sliding, (ii) Rolling, (iii) Crumpling, (iv) Winding.

It is also necessary to assume that in each case, where applicable, the action takes place over a constant horizontal width a of snow. This, from observation, is approximately true.

(4) In what follows, m is the mass of moving snow; g is the acceleration of gravity; r, for case (ii) and (iv), is the radius of the rolling body; x is the distance moved by the body's center of mass, down the slope, in time t; μ, for case (i) and (iii), is the coefficient of friction snow/slope; and θ, for case (ii) and (iv), is the angle turned through by the body in time t.

(5) (i) *Sliding*. The acceleration is well known to be $g(\sin\alpha-\mu\cos\alpha)$.

(ii) *Rolling*. If the body is a uniform cylinder, axis horizontal, the acceleration is well known to be $\frac{2}{3}g\sin\alpha$. Accelerations for bodies of other shapes, axi-symmetrical about a horizontal axis, can be calculated without difficulty (e.g. for a sphere it is $\frac{5}{7}g\sin\alpha$). They vary between the extreme cases: (1) all mass concentrated on the axis: $g\sin\alpha$, and (2) hollow cylinder: $\frac{1}{2}g\sin\alpha$. We may say that for the average-shaped rock the acceleration is roughly $\frac{3}{4}g\sin\alpha$. Since we have assumed incompressibility of the snow, there is no rolling friction.

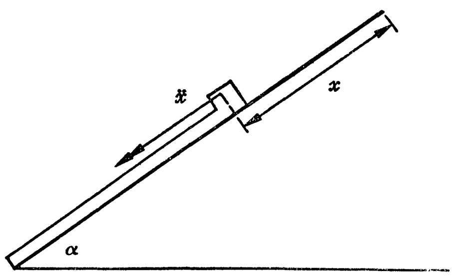


FIG. 1

(iii) *Crumpling* (Fig. 1). At time t, $m=\rho abx$ and the force acting down the slope is $mg(\sin\alpha-\mu\cos\alpha)=mgp$, say. (If the distance of the centre of mass from the slope increases, the normal reaction will be $>mg\cos\alpha$. However the effect is slight and will be ignored.)

So $mgp=(d/dt)$ (momentum), i.e., $\rho abxgp=(d/dt)(\rho abx\dot{x})$, giving $x\ddot{x}+\dot{x}^2-pgx=0$, with solution (obtained by writing $\dot{x}^2=q$, or by trial with $x=st^u$) to suit starting conditions: $x=\frac{1}{6}pgt^2$, so that

$$\ddot{x}=\frac{g}{3}(\sin\alpha-\mu\cos\alpha).$$

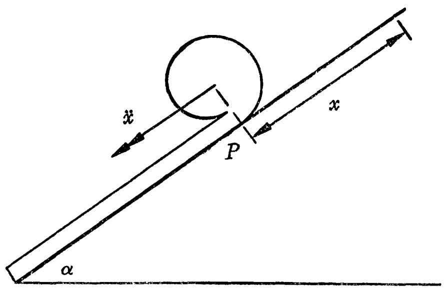


FIG. 2

(iv) *Winding* (Fig. 2). For large θ we may treat the rolling body as a cylinder and $r = b\theta/2\pi$, so that $m = \rho ab^2\theta^2/4\pi$.

About the horizontal line through the instantaneous centre P, the moment of inertia is

$$I = \frac{3}{2} mr^2 = \frac{3\rho ab^4\theta^4}{32\pi^3},$$

and the acting torque about this line is G, where

$$G = mgr \sin \alpha = \frac{\rho ab^3 g\theta^3 \sin \alpha}{8\pi^2}.$$

But $G = (d/dt)\ [I\dot{\theta}]$, so simplifying,

$$\left\{\frac{4\pi g \sin \alpha}{3b}\right\} \theta^3 = \frac{d}{dt} (\theta^4\dot{\theta}).$$

If $\{\ \} = v^2$, we get, dividing by θ^3, $\theta\ddot{\theta} + 4\dot{\theta}^2 - v^2 = 0$ with solution (obtained as before) to suit starting conditions:

$$\theta = \tfrac{1}{2}vt,$$

so that $\dot{x} = r\dot{\theta}$, which on substitution reduces to

$$\dot{x} = \frac{gt}{6} \sin \alpha, \quad \text{giving} \quad \ddot{x} = \frac{g}{6} \sin \alpha.$$

(6) In the above calculations, μ is the coefficient of kinetic friction. As every skier knows, this is small, in fact it is of a different order from that of the coefficient of static friction. The effect is due to the Clausius-Clapeyron equation as applied to ice under pressure. Indeed, the sport of skiing is made possible by this phenomenon. Therefore, as a first approximation, we may ignore μ in our results. If we put $g' = g \sin \alpha$, so that g' is the acceleration of a particle on the smooth mountain slope, our results for acceleration can be tabulated as follows:

	(i) Sliding,	(ii) Rolling,	(iii) Crumpling,	(iv) Winding.
Accn:	g'	$\frac{3}{4}g'$	$\frac{1}{3}g'$	$\frac{1}{6}g'$

and the accelerations decrease in the order given.

(7) The fact that rocks and hard lumps of snow (rolling type) accelerate faster than crumpling or winding avalanches is borne out, to a certain extent, by observation: such objects appear to go ahead of the main avalanche, although, in the cloud of fine snow that surrounds an avalanche, it is difficult to see precisely what is going on. In any case, anyone watching an avalanche, if on a snow slope himself, is not in the right frame of mind to make detached scientific observations!

THE SPECTRAL THEOREM FOR REAL SYMMETRIC MATRICES

P. C. SHIELDS, Wayne State University

The purpose of this note is to provide a proof of the

LEMMA. *If T is a symmetric linear operator on a finite dimensional real Hilbert space, then T has a **real** eigenvalue.*

The spectral theorem for real, symmetric matrices is a simple consequence of this lemma. One standard method for proving the lemma is to detour through the analogous result for hermitian operators. A second method is to show that the quadratic form (Tx, Tx) has a nonnegative maximum λ^2 on the unit sphere and then that either λ or $-\lambda$ is an eigenvalue. The proof given here is quite different and is a generalization of the following 2×2 proof:

The characteristic polynomial of

$$A = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \quad \text{is} \quad \lambda^2 - (a+c)\lambda + ac - b^2.$$

The existence of a real eigenvalue follows because the discriminant is nonnegative:

$$(a+c)^2 - 4(ac - b^2) = (a-c)^2 + 4b^2.$$

Proceeding to the proof of the lemma, let p be the minimum polynomial of T, and suppose p has an irreducible real quadratic factor, that is,

$$p(\lambda) = g(\lambda)h(\lambda), \qquad \text{where } g(\lambda) = \lambda^2 + b\lambda + c$$

with

$$b^2 - 4c < 0. \tag{1}$$

Put $M = \{x \mid g(T)x = 0\}$ and note that

(2) M is a nontrivial invariant subspace and g is the minimum polynomial for the restriction T/M of T to M.

Since p doesn't divide h there must be a vector x such that $h(T)x \neq 0$. The vector $h(T)x$ is then a nonzero vector in M. The invariance of M is trivial. If q is the minimum polynomial for T/M, then q must divide g for $g(T)x=0$ for $x \in M$. The irreducibility of g gives $g=q$ and completes the proof of (2).

Choose $x \in M$ such that $Tx \neq 0$. We have

$$0 = g(T)x = T^2x + bTx + cx$$

so that $0=(g(T)x, x)=(Tx, Tx)+b(Tx, x)+c(x, x)$ and hence

$$(|Tx|^2 + c|x|^2)^2 = b^2(Tx, x)^2.$$

The Schwarz inequality, combined with (1) and $Tx \neq 0$, implies

$$(|Tx|^2 + c|x|^2)^2 \leqq b^2|Tx|^2|x|^2 < 4c|Tx|^2|x|^2,$$

so we arrive at the false statement $(|Tx|^2 - c|x|^2)^2 < 0$.

This contradiction establishes that the minimum polynomial p of a symmetric operator T cannot have irreducible quadratic factors and completes the proof of the lemma. This proof is similar to that given in the author's *Linear Algebra*, Addison-Wesley, 1963.

Query: The Perron-Frobenius Theorem has a simple 2×2 proof. Can one use something like the above methods to obtain the general result?

A GENERALIZATION OF NANJUNDIAH'S IDENTITY

M. T. L. Bizley, London, England

In 1958 Nanjundiah (this Monthly, page 354) published a combinatorial identity which, with a change of notation, may be written:

$$\sum_i \binom{b}{i}\binom{c}{i}\binom{a+i}{b+c} = \binom{a}{b}\binom{a}{c}, \tag{1}$$

the sum being over all integral values of i for which the expression summed does not vanish.

The purpose of this note is to state and prove the two identities:

$$\begin{cases} \sum_i \binom{b}{i}\binom{c}{i-d}\binom{a+i}{b+c} = \binom{a}{b-d}\binom{a+d}{c+d} & (2) \\ \sum_i \binom{b}{i}\binom{c}{d-i}\binom{a+i}{b+c} = \binom{a}{b+c-d}\binom{a-c+d}{d}. & (3) \end{cases}$$

It will be seen that Nanjundiah's identity (1) is a special case of (2), namely when $d=0$; also of (3), namely when $d=c$.

To prove (2) and (3), we begin with the obvious relation:

$$(1+x)^m(1+y)^n[1+x(1+y)]^b = (1+x)^m(1+y)^n[(1+x)+xy]^b. \tag{4}$$

Expanding each square bracket as a binomial in the particular form shown, and equating coefficients of $x^p y^q$ (for arbitrary p, q) on the two sides of (4), we find:

$$\sum_i \binom{b}{i}\binom{m}{p-i}\binom{n+i}{q} = \sum_i \binom{b}{i}\binom{m+i}{p-b+i}\binom{n}{q-b+i}, \text{ i.e.,}$$

$$\sum_i \binom{b}{i}\binom{m}{p-i}\binom{n+i}{q} = \sum_i \binom{b}{i}\binom{m+i}{m-p+b}\binom{n}{q-b+i}. \tag{5}$$

This is true for all positive integers b, m, n, p, q, with the usual convention that $\binom{u}{v}=0$ if $u<v$ or $v<0$. As it stands, (5) is not interesting because each side is a sum. However, by assigning suitable values to p and q, we can reduce one side of (5) to a single term, as we now show.

To derive (2), we put $m=a$; $n=c$; $p=a-c$; $q=b-d$. Then (5) becomes:

$$\sum_i \binom{b}{i}\binom{a}{a-c-i}\binom{c+i}{b-d} = \sum_i \binom{b}{i}\binom{a+i}{b+c}\binom{c}{i-d}. \tag{6}$$

But since

$$\binom{a}{a-c-i}\binom{c+i}{b-d} = \binom{a}{b-d}\binom{a-b+d}{a-c-i}$$

identically, the left side of (6) is

$$\binom{a}{b-d}\sum_i \binom{b}{i}\binom{a-b+d}{a-c-i} = \binom{a}{b-d}\binom{a+d}{a-c}$$
$$= \binom{a}{b-d}\binom{a+d}{c+d}.$$

Hence,

$$\sum_i \binom{b}{i}\binom{c}{i-d}\binom{a+i}{b+c} = \binom{a}{b-d}\binom{a+d}{c+d},$$

which is (2). Again, to derive (3), we return to (5) and put $m=c$; $n=a$; $p=d$; $q=b+c$. Then (5) becomes:

$$\sum_i \binom{b}{i}\binom{c}{d-i}\binom{a+i}{b+c} = \sum_i \binom{b}{i}\binom{c+i}{c-d+b}\binom{a}{c+i}. \tag{7}$$

But since

$$\binom{c+i}{c-d+b}\binom{a}{c+i} = \binom{a}{b+c-d}\binom{a-b-c+d}{a-c-i}$$

identically, the right side of (7) is

$$\binom{a}{b+c-d}\sum_i \binom{b}{i}\binom{a-b-c+d}{a-c-i} = \binom{a}{b+c-d}\binom{a-c+d}{a-c}$$
$$= \binom{a}{b+c-d}\binom{a-c+d}{d}.$$

Hence,

$$\sum_i \binom{b}{i}\binom{c}{d-i}\binom{a+i}{b+c} = \binom{a}{b+c-d}\binom{a-c+d}{d},$$

which is (3).

CONVERGENCE OF PICARD'S METHOD FOR $|\lambda|>|\lambda_1|$

J. W. Burgmeier, University of Vermont,
and M. R. Scott, Sandia Laboratories

In many textbooks [1, p. 142; 2, p. 263; 3, p. 281] it is stated that the radius of convergence of Picard's method for a nonhomogeneous integral equation is equal to the absolute value of the smallest characteristic value, while others [4] simply state that the iteration method converges uniformly for $|\lambda|<|\lambda_1|$ without mentioning what happens when $|\lambda|>|\lambda_1|$. (In order to distinguish between eigenvalues and reciprocal eigenvalues, we suggest, as have others [5, 6], the following: A **characteristic value** is a scalar λ such that

$$\phi(x) = \lambda\int_0^1 K(x, y)\phi(y)dy$$

possesses a nontrivial solution ϕ; i.e., λ is the reciprocal of an eigenvalue.) We shall prove that, under certain circumstances, the radius of convergence can be much larger than $|\lambda_1|$.

Let us consider an integral equation of the form

$$u(x) = f(x) + \lambda\int_0^1 K(x, y)u(y)dy, \tag{1}$$

where $K(x, y)$ is a symmetric L_2-kernel on the rectangle $0\leqq x,\ y\leqq 1$ and $f(x)\in L_2(0, 1)$. We denote the eigenfunctions and characteristic values (counting multiplicities) of the homogeneous equation (i.e., when $f(x)\equiv 0$) by $\phi_n(x)$ and λ_n, respectively. In addition, we let

$$\psi(x;\lambda) = f(x) + \sum_{n=1}^{\infty}\lambda^n K^{(n)}f(x), \tag{2}$$

where $K^{(n)}f(x)=\int_a^b K^{(n)}(x, y)f(y)dy$, denote the Neumann series for (1), which is generated by the usual Picard successive approximation process with $f(x)$ as the initial approximation.

THEOREM. *Let* $K(x, y)$ *be such that*

$$K(x, y) = \sum_{j=1}^{\infty} \frac{\phi_j(x)\phi_j(y)}{\lambda_j} \tag{3}$$

converges uniformly and absolutely on $0 \leqq x, y \leqq 1$. *Let the function* $f(x)$ *be such that* $(f, \phi_j) = 0$, $j = 1, 2, \cdots, k-1$. *Then the Neumann series* (2) *converges for all* $|\lambda| < |\lambda_k|$, *and*

(a) *if* $\lambda \notin \{\lambda_1, \lambda_2, \cdots, \lambda_{k-1}\}$, *then* $\psi(x; \lambda)$ *is the unique solution to* (1),

(b) *if* $\lambda = \lambda_j$, $1 \leqq j \leqq k-1$, *then* $\psi(x; \lambda)$ *is the unique solution to* (1) *which is also orthogonal to all of the eigenfunctions corresponding to* λ_j.

Proof: Since (3) converges absolutely and uniformly, so does

$$K^{(n)}(x, y) = \sum_{j=1}^{\infty} \frac{\phi_j(x)\phi_j(y)}{\lambda_j^n}, \quad n = 1, 2, \cdots.$$

Thus

$$K^{(n)}f(x) = \sum_{j=1}^{\infty} \frac{\phi_j(x)}{\lambda_j^n} \int_0^1 \phi_j(y)f(y)dy = \sum_{j=k}^{\infty} \frac{(f, \phi_j)}{\lambda_j^n} \phi_j(x),$$

where we have used the uniform convergence and $f \perp \phi_1, \phi_2, \cdots, \phi_{k-1}$. Hence

$$\sum_{n=1}^{\infty} \lambda^n K^{(n)}f(x) = \sum_{n=1}^{\infty} \lambda^n \sum_{j=k}^{\infty} \frac{(f, \phi_j)}{\lambda_j^n} \phi_j(x) = \sum_{j=k}^{\infty} (f, \phi_j)\phi_j(x) \sum_{n=1}^{\infty} \left(\frac{\lambda}{\lambda_j}\right)^n.$$

For $|\lambda| < |\lambda_k| \leqq |\lambda_j|$, $j \geqq k$, we can sum the second series to obtain

$$\psi(x; \lambda) - f(x) = \sum_{n=1}^{\infty} \lambda^n K^{(n)}f(x) = \sum_{j=k}^{\infty} \frac{\lambda(f, \phi_j)}{\lambda_j - \lambda} \phi_j(x). \tag{4}$$

From the uniform convergence and the condition $f \perp \phi_1, \phi_2, \cdots, \phi_{k-1}$, it is straightforward to show that $\psi(x; \lambda)$ does satisfy (1). Also for

$$\lambda \notin \{\lambda_1, \lambda_2, \cdots, \lambda_{k-1}\},$$

the solution is unique [1, p. 116].

When $\lambda = \lambda_j$, $1 \leqq j \leqq k-1$, the conditions $f \perp \phi_1, \phi_2, \cdots, \phi_{k-1}$ insure that there does exist a solution to (1). However, it is not unique. But, by requiring that the solution ψ be orthogonal to all of the eigenfunctions of λ_j, we do obtain a unique solution [1, p. 116]. For $i = 1, 2, \cdots, k-1$, we have

$$(\psi, \phi_i) = (f, \phi_i) + \lambda \sum_{j=k}^{\infty} \frac{(f, \phi_j)}{\lambda_j - \lambda} (\phi_j, \phi_i) = 0.$$

Thus the theorem is proved.

We now mention a few consequences of the above theorem. If K is non-degenerate (i.e., possessing an infinite number of eigenvalues) and if the eigen-

functions form a complete orthonormal system in $L_2(0, 1)$, then there does *not* exist an f such that $(f, \phi_j)=0$ for all j, except $f\equiv 0$. As a consequence, the Picard method *cannot* converge for all λ in this case.

On the other hand, for degenerate kernels and nondegenerate kernels whose eigenfunctions do not form a complete orthonormal system in $L_2(0, 1)$, there exists a large class of functions f which are orthogonal to all of the eigenfunctions and, for such f, Picard's method converges for all λ. In fact, the convergence is trivial, since any function which is orthogonal to all of the eigenfunctions of a symmetric kernel is also orthogonal to the kernel itself [**4**, p. 106]. Thus $\psi\equiv f$.

A portion of this work was supported by the U. S. Atomic Energy Commission.

References

1. R. Courant and D. Hilbert, Methods of Mathematical Physics, Vol. 1, Interscience, New York, 1953.

2. Francis B. Hildebrand, Methods of Applied Mathematics, Prentice Hall, London, 1965.

3. S. G. Mikhlin and K. L. Smolitskiy, Approximate Methods for Solution of Differential and Integral Equations, American Elsevier, New York, 1967.

4. F. G. Tricomi, Integral Equations, Interscience, New York, 1963.

5. H. P. Thielman, Applications of the Fixed Point Theorem, *Nonlinear Integral Equations*, University of Wisconsin Press, 1964, pp. 35–48.

6. R. B. Rall, Variational Methods for Nonlinear Integral Equations, *Nonlinear Integral Equations*, University of Wisconsin Press, 1964, pp. 155–189.

RESEARCH PROBLEMS

EDITED BY VICTOR KLEE

In this Department the Monthly presents easily stated research problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Victor Klee, Department of Mathematics, University of Washington, Seattle, WA 98105.

CAN EVERY TRIANGLE BE DIVIDED INTO n TRIANGLES SIMILAR TO IT?

R. W. FREESE, ANN K. MILLER, Saint Louis University, and
ZALMAN USISKIN, University of Chicago

DEFINITION. Given any triangle ABC and natural number n, then n is an acceptable number for triangle ABC if there exists a decomposition of the triangle into n subtriangles, with each subtriangle similar to triangle ABC. Thus, for every triangle, 1 is an acceptable number. Further, given any triangle ABC with M the midpoint of side AB, N the midpoint of side BC, and P the midpoint of AC, then the subdivision of ABC into subtriangles AMP, BMN, CNP, and MNP yields four triangles each similar to the original one. Thus, for every

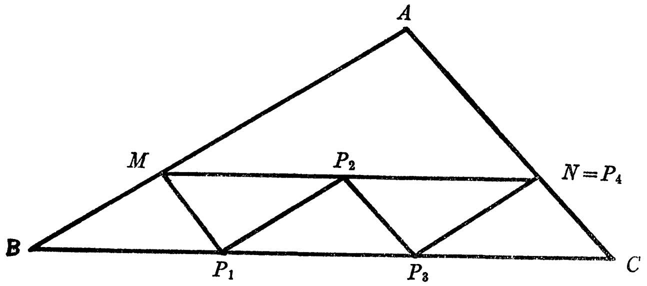


FIG. 1

triangle, 1, 4, 7, $\cdots$ are acceptable numbers. Again, let ABC be any triangle with base BC. On side AB mark off a point M such that the length of the segment MB is one-third of the length of side AB. Construct the line through M parallel to side BC and let N be the point on side AC cut by the parallel line. Next construct the following sequence of points: let P_1 be the point on side BC cut by the line through M parallel to side AC, let P_2 be the point of intersection of MN and the line through P_1 parallel to AB, etc. Then it can be shown that P_4 coincides with N and that each resulting subtriangle is similar to the original one. Therefore, 6, 9, 12, $\cdots$ are acceptable numbers for every triangle. In an analogous fashion, if the point M is chosen on side AB such that the length of MB is one-fourth of the length of AB and if the points P_1 to P_6 are constructed in a manner similar to the preceding method, then it can be shown that 8, and hence 11, 14, 17, $\cdots$ are acceptable for every triangle. Thus, we have shown the following:

THEOREM. *For any natural number n, $n \neq 2$, 3 or 5, there exists a decomposition of any triangle into n subtriangles similar to it.*

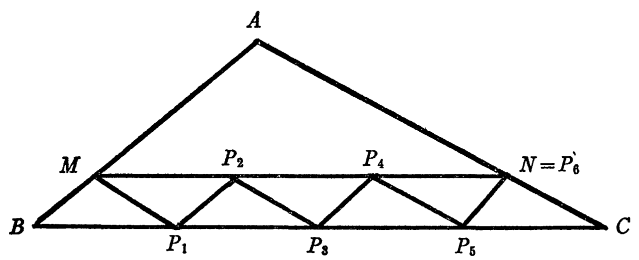
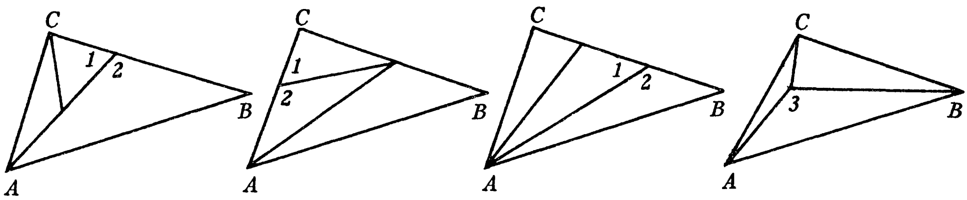


FIG. 2

Consider now, the special case of a right triangle. Let ABC be any such triangle, with the angle at A the right angle. If AM is the altitude to base BC, then the decomposition of triangle ABC into subtriangles ABM and ACM yields two triangles similar to the original triangle. Therefore, for right triangles, every

natural number is acceptable. On the other hand, if a triangle has 2 as an acceptable number, then since the decomposition must be accomplished by joining a vertex to an opposite side, an inspection of the cases reveals that the triangle must be a right triangle. Consider next the case $n=3$. We first note that there are only 4 essentially different ways of dividing a triangle into 3 triangles. (The first three are the only ways if one side of a smaller triangle joins a vertex to the opposite side; the latter is the only other possibility.) They are drawn here.



(a) (b) (c) (d)

In cases (a), (b), and (c), $m\angle 1+m\angle 2=180°$. If the smaller triangles are to be similar to $\Delta\, ABC$, then angles of $\Delta\, ABC$ must be congruent to angles 1 and 2. If angles 1 and 2 are congruent to different angles of $\Delta\, ABC$, then two angles of the $\Delta\, ABC$ add to 180—impossible. So angles 1 and 2 must be congruent to the same angle of $\Delta\, ABC$, so $\angle 1 \cong \angle 2$, in which case they are right angles and $\Delta\, ABC$ is a right triangle.

In case (d), suppose angle C (without loss of generality) is the largest angle of $\Delta\, ABC$. It is well known that $\angle 3 > \angle C$. Hence the triangle with $\angle 3$ in it could not be similar to the original triangle.

Hence, only a right triangle can be divided into three triangles, all similar to each other and to it.

These preliminaries lead naturally into the following questions: Does the set of acceptable numbers of a triangle determine its shape; e.g. if a triangle has 5 as acceptable, must it be a right triangle? Conversely, given the shape of a triangle, what is the set of acceptable numbers for it?

A PROBLEM OF ILLUMINATION

L. Fejes Tóth, Hungarian Academy of Sciences

Along an infinite straight road we want to erect equal lamps with a given average number per mile. How should the lamps be distributed so as to maximize the illumination of the road at the worst illuminated point?

More precisely, let $f(x)$ be the illumination function defined for $x \geqq 0$ and giving the illumination of the road due to one lamp at a distance x from its foot. The function $f(x)$ is supposed to be continuous, nonnegative, nonincreasing and such that

$$\int_0^\infty f(x)dx < \infty.$$

If $\cdots < x_{-2} < x_{-1} < x_0 < x_1 < x_2 \cdots$ are the coordinates of the lamps then the (upper) density D of the lamps is defined by

$$D = \overline{\lim_{R\to\infty}}\ N_R/2R,$$

where N_R is the number of the lamps in the interval $(-R, R)$. If D is finite then so is the illumination of the road $I(x) = \sum_i f(|x - x_i|)$ at any point x. We want to find

$$M(f, D) = \sup \inf_x I(x)$$

for all possible arrangements of the lamps with the prescribed density D.

This problem was raised by the author at the Colloquium on Discrete Geometry in Oberwolfach in 1962. L. Danzer soon observed that it is not always the point-lattice (equidistant distribution) which gives the solution of the problem. The following example is due to A. Heppes [*Egy egydimenzios probléma*, Mat. Lapok, 14 (1963) 124–127]: Let $f(x)$ be defined by

$$f(x) = \begin{cases} 1 - \frac{1}{2}x, & 0 \leqq x < 1 \\ \frac{3}{2} - x, & 1 \leqq x < \frac{3}{2} \\ 0, & \frac{3}{2} \leqq x. \end{cases}$$

If the lamps are equally spaced with $D = 1$ then $\min_x I(x) = 3/2$. On the other hand, placing the lamps at the points $\pm\frac{1}{4}$, $\pm 2 \pm \frac{1}{4}$, $\pm 4 \pm \frac{1}{4}$, $\cdots$ we obtain a constant illumination $I(x) = \frac{7}{4} > \frac{3}{2}$.

Since the determination of $M(f, D)$ for any f and D, by a formula or any other sort of general instruction, seems to be extremely difficult, we give our problem the following definitive form: Prove or disprove the conjecture that for any f and D, $M(f, D)$ can be attained by a distribution consisting of the union of a finite number of congruent point-lattices.

CLASSROOM NOTES

EDITED BY DAVID DRASIN

Manuscripts for this Department should be sent to David Drasin, Division of Mathematical Sciences, Purdue University, Lafayette, IN 47907.

A NEW PROOF OF THE EXISTENCE OF THE FREE GROUP

N. C. MEYER, JR., The University of Oregon

The following is a new and simpler proof of the existence of a free group on an arbitrary set.

DEFINITION. *Let A be a set. A free group on A is a group F and a function $f: A \to F$, such that if G is any group and $g: A \to G$ a function, then there is a unique homomorphism $h: F \to G$ such that $g = hf$.*

Intuitively, F consists of the elements of A, formal symbols a^{-1} where $a \in A$, and all finite chains ("words") of a's and a^{-1}'s. One defines multiplication to be juxtaposition (defined below), then one "reduces" words by cancelling out any occurrence of aa^{-1} or $a^{-1}a$. The usual proofs that F is a group (especially showing associativity) involve a lengthy case by case argument. The unreduced words form a semi-group (a set with an associative operation) upon which we wish to define an equivalence relation, reduction; in the present proof, we avoid the consideration of cases by defining the equivalence relation in a new way. First we need two lemmas concerning semigroups.

LEMMA 1. *Let A be a nonempty set; let X be the set of all finite chains $(a_1\, a_2 \cdots a_n)$, where each $a_i \in A$; define an operation on X by: $(a_1 \cdots a_n)(b_1 \cdots b_m) = (a_1 \cdots a_n b_1 \cdots b_m)$. This operation makes X a semigroup (called the* free semigroup*); define $j: A \to X$ by $j(a) = (a)$, the chain of length one. If Y is any semigroup and $k: A \to Y$ a function, then there is a unique homomorphism $f: X \to Y$ such that $k = fj$.*

Proof. Given any a_i's, b_i's, and c_i's in A,

$$\begin{aligned}[(a_1 \cdots a_n)(b_1 \cdots b_m)](c_1 \cdots c_p) &= (a_1 \cdots a_n b_1 \cdots b_m)(c_1 \cdots c_p)\\ &= (a_1 \cdots a_n b_1 \cdots b_m c_1 \cdots c_p)\\ &= (a_1 \cdots a_n)[(b_1 \cdots b_m)(c_1 \cdots c_p)].\end{aligned}$$

So X is a semigroup. If $k: A \to Y$, define $f: X \to Y$ by $f((a_1 \cdots a_n)) = k(a_1) \cdots k(a_n)$. Then f is a homomorphism since

$$f((a_1 \cdots a_n)(b_1 \cdots b_m)) = [k(a_1) \cdots][k(b_1) \cdots],$$

and $fj = k$. If $gj = k$ also, then $g(a_1 \cdots a_n) = g((a_1) \cdots (a_n)) = k(a_1) \cdots k(a_n) = f(a_1 \cdots a_n)$, so f is unique.

LEMMA 2. *If Y is a semigroup and $*$ is an equivalence relation on Y such that $x*y$ and $z*w$ imply $xz*yw$, then $Y/*$ is a semigroup, and $p: Y \to Y/*$ is a homomorphism. If $h: Y \to Z$ is a semigroup homomorphism such that $x*y$ implies $h(x) = h(y)$, then there is a unique homomorphism $\bar{h}: Y/* \to Z$ such that $\bar{h}p = h$.*

Proof. Let $\bar{x} = p(x) =$ the equivalence class of x. Define $\bar{x}\bar{y} = \overline{xy}$. If $x*y$ and $z*w$ then $\overline{xz} = \overline{yw}$ since $xz*yw$, so this operation is well defined. Next $(\bar{x}\bar{y})\bar{z} = \overline{xy}\bar{z} = \overline{xyz} = \bar{x}(\overline{yz})$, so $Y/*$ is a semigroup. Now $p(xy) = \overline{xy} = \bar{x}\bar{y} = p(x)p(y)$, so p is a homomorphism. If $h: Y \to Z$ such that $x*y$ implies $h(x) = h(y)$, define $\bar{h}$ by $\bar{h}(\bar{x}) = h(x)$. h is well defined since $\bar{x} = \bar{y}$ implies $x*y$, which implies $h(x) = h(y)$.

$$\bar{h}(\bar{x}\bar{y}) = \bar{h}(\overline{xy}) = h(xy) = h(x)h(y) = \bar{h}(\bar{x})\bar{h}(\bar{y})$$

so $\bar{h}$ is a homomorphism. Finally if $f: Y/* \to Z$ with $fp = h$, then

$$f(\bar{x}) = f(p(x)) = h(x) = \bar{h}(\bar{x}), \quad \text{so } f = \bar{h}.$$

THEOREM. *If A is any nonempty set, there is a free group on A.*

Proof. Let $A' = \{a' \mid a \text{ in } A\}$, where we assume $A \cap A' = \varnothing$ (to be strictly correct replace A by $\{(a, 1) \mid a \text{ in } A\}$, and A' by $\{(a, -1) \mid a \text{ in } A\}$). Let X be the free semigroup on the set $A \cup A'$ as defined in Lemma 1, with $j: A \cup A' \to X$. We write $j(a) = a$ and $j(a') = a'$.

Define the relation $*$ on X by: $x * y$ iff $g(x) = g(y)$ for all semigroup homomorphisms $g: X \to G$ satisfying:

$$(\#) \qquad G \text{ is a group, and } g(a') = g(a)^{-1} \text{ for all } a \text{ in } A.$$

$*$ is an equivalence relation since for all g satisfying #, and all x, y, z in X,

$$g(x) = g(x);$$
$$g(x) = g(y) \quad \text{implies } g(y) = g(x);$$
$$g(x) = g(y) \quad \text{and} \quad g(y) = g(z) \quad \text{imply } g(x) = g(z).$$

If $x*y$ and $z*w$, then for all g satisfying #, $g(xz) = g(x)g(z) = g(y)g(w) = g(yw)$, so $xz*yw$. Thus, by Lemma 2, $X/*$ is a semigroup and $p: X \to X/*$ is a homomorphism. We shall show that $X/*$ is a group and that $pj|_A: A \to X/*$ is a free group on A. Pick a in A, let x be in X, then for any g satisfying #, $g(aa') = 1$, so $g(xaa') = g(x)$. So $x * xaa'$ or $\bar{x} = \bar{x}(\overline{aa'})$, and similarly $\bar{x} = (\overline{aa'})\bar{x}$. Therefore $\overline{aa'}$ is the identity of $X/*$. Now if x is in X, $x = (b_1 \cdots b_n)$, where each b_i is in $A \cup A'$. For a in A, denote $a'' = a$; let $y = (b_n' \cdots b_1')$. For g satisfying #,

$$\begin{aligned} g(xy) &= g(b_1) \cdots g(b_n)g(b_n') \cdots g(b_1') = g(b_1) \cdots g(b_n)g(b_n)^{-1} \cdots g(b_1)^{-1} \\ &= g(x)g(x)^{-1} = 1 = g(aa') \end{aligned}$$

so $xy * aa'$, i.e., $\bar{x}\bar{y} = 1$ in $X/*$. Similarly, $\bar{y}\bar{x} = 1$, and so $X/*$ is a group. Note that $p(a') = p(a)^{-1}$.

Now let $k: A \to G$, where G is a group; extend k to $k': A \cup A' \to G$ by $k'(a') = k(a)^{-1}$ for a in A. By Lemma 1, there is a unique homomorphism $h: X \to G$ such that $hj = k'$. Note that

$$h(a') = h(j(a')) = k'(a') = k'(a)^{-1} = h(a)^{-1}$$

and so h satisfies #, and therefore $x*y$ implies $h(x) = h(y)$. By Lemma 2 there is a unique homomorphism $\bar{h}: X/* \to G$ so that $\bar{h}p = h$ (since $X/*$ and G are groups,

$$\begin{tikzcd}
X/* \arrow[ddddr, "\bar{h}"] & \\
X \arrow[u, "p"] \arrow[dddr, "h"] & \\
A \cup A' \arrow[u, "j"] \arrow[ddr, "k'"] & \\
\cup & \\
A \arrow[r, "k"] & G
\end{tikzcd}$$

$\bar{h}$ is a group homomorphism). Thus $\bar{h}(pj) = hj = k'$ and $\bar{h}(pj|_A) = k$. Furthermore $\bar{h}$ is unique since if $f(pj|_A) = k$, then $(fp)j = k'$ since

$$fpj(a') = f(p(a)^{-1}) = (fpj(a))^{-1} = k(a)^{-1} = k'(a');$$

and so by Lemma 1, $fp = h$, and by Lemma 2, $f = \bar{h}$. Therefore, $pj|_A: A \to X/*$ is a free group on the set A.

REMARK: The free product (coproduct) of a set $\{G_\alpha\}$ of groups can also be obtained in this way. Let $j: \bigcup_\alpha G_\alpha \to X$ be the free semigroup on $\bigcup_\alpha G_\alpha$ and $j_\alpha = j|_{G_\alpha}: G_\alpha \to X$. Define $x * y$ in X iff $h(x) = h(y)$ for all homomorphisms $h: X \to H$, where H is a group and $hj_\alpha: G_\alpha \to H$ is a homomorphism for all α. As above, one checks that $X/*$ is a group, that $pj_\alpha: G_\alpha \to X/*$ is a monomorphism, and that $(X/*, \{pj_\alpha\})$ satisfies the appropriate universal diagram.

References

1. Emil Artin, The free products of groups, Amer. J. Math., 69 (1947) 1–4.
2. C. Chevalley, Fundamental Concepts of Algebra, Academic Press, New York, 1956, pp. 40–43.
3. M. Hall, The Theory of Groups, Macmillan, New York, 1959, pp. 91–93.
4. A. G. Kurosh, The Theory of Groups, Chelsea, New York, 1960, pp. 124–29.
5. S. Lang, Algebra, Addison-Wesley, Reading, Mass., 1965, pp. 33–36.
6. J. Rotman, The Theory of Groups, An Introduction, Allyn and Bacon, Boston, 1965, pp. 237–239.
7. W. R. Scott, Group Theory, Prentice-Hall, Englewood Cliffs, N. J., 1964, pp. 174–183.

IDEMPOTENCE IN CARDINAL ARITHMETIC

R. D. KOPPERMAN, City College of New York

We present a proof of idempotence in cardinal multiplication which is shorter than existing proofs but uses more properties of ordinal arithmetic.

Suppose ordinals have been defined as well as cardinals (=initial ordinals) and we have defined $<$, $\leqq$, $=$ for ordinals, $\approx$ (equipollence), $\preccurlyeq$, and $\prec$. We also assume it known that for any two ordinals, $\alpha \leqq \beta$ or $\beta < \alpha$, and $\alpha \leqq \beta$ if and only if $\alpha \subset \beta$ (from which we obtain the fact that for any two ordinals, $\alpha \preccurlyeq \beta$ or $\beta \prec \alpha$), and we use the Schröder-Bernstein Theorem.

Let X be a well-ordered set, Y_x well-ordered for each $x \in X$. Then by $S_{x\in X} Y_x$ we mean $\bigcup_{x\in X} Y_x \times \{x\}$, well-ordered by $\langle y, x\rangle < \langle y', x'\rangle$ iff $x < x'$ (in the given well-ordering of X) or $x = x'$ and $y < y'$ (in the given well-ordering of Y_x). It is well known and easy to check that the above actually defines a well-ordering. If α is an ordinal and for $\gamma \in \alpha$, β_γ is an ordinal, then by $\sum_{\gamma<\alpha} \beta_\gamma$ we mean the ordinal order-isomorphic to $S_{\gamma\in\alpha}\beta_\gamma$. If α and β are ordinals, then $\alpha \cdot \beta = \sum_{\gamma<\alpha}\beta_\gamma$ where $\beta_\gamma = \beta$ for each $\gamma < \alpha$.

From the above definition it is clear that

$$\sum_{\gamma<\alpha} \beta_\gamma \approx \bigcup_{\gamma<\alpha} \beta_\gamma \times \{\gamma\}$$

and therefore $\alpha\beta \approx \alpha\times\beta$, $\alpha+\alpha\approx\alpha\times 2$. Also clear is the fact that if α is a limit ordinal,

$$\sum_{\gamma<\alpha}\beta_\gamma = \bigcup_{\gamma<\alpha}\sum_{\delta<\gamma}\beta_\delta.$$

If A is well-ordered and $\{C_a \mid a\in A\}$ a collection of sets, then

$$\bigcup_{a\in A} C_a \preccurlyeq \bigcup_{a\in A} C_a\times\{a\}$$

by the map $f(y)=\langle y, b\rangle$ for b the first element of A such that $y\in C_b$. Finally note that in the proof to follow we do not assume it known that all infinite cardinals are limit ordinals; if this fact were assumed, the second paragraph of the proof might be eliminated.

THEOREM. *If α is an infinite ordinal then $\alpha\approx\alpha^2$.*

Proof: We have $\alpha\approx\alpha\times 1\preccurlyeq\alpha\times\alpha\approx\alpha^2$ so if our theorem is false there is an infinite ordinal, thus a least infinite ordinal α such that $\alpha\prec\alpha^2$. Our α must be a cardinal, for if $\beta<\alpha$, $\beta\approx\alpha$, then

$$\alpha^2\approx\beta^2\approx\beta\approx\alpha.$$

This α cannot be a successor, for if $\beta+1=\alpha$, then $\beta<\alpha$,so:

$$\beta\preccurlyeq\beta+1\preccurlyeq\beta+\beta\approx\beta\times 2\preccurlyeq\beta\times\beta\approx\beta^2\approx\beta,$$

thus $\alpha\approx\beta$, contradicting the fact that α is a cardinal.

However if α is a cardinal and a limit ordinal, then

$$\alpha\times\alpha = \bigcup_{\beta<\alpha}\beta\times\beta\preccurlyeq\bigcup_{\beta<\alpha}\beta\times\beta\times\{\beta\}\approx\sum_{\beta<\alpha}\beta^2 = \bigcup_{\gamma<\alpha}\left(\sum_{\beta<\gamma}\beta^2\right).$$

Thus since $\alpha\prec\alpha^2$, $\alpha\prec\sum_{\beta<\alpha}\beta^2$, so $\alpha<\sum_{\beta<\alpha}\beta^2$, i.e., for some $\gamma<\alpha$

$$\alpha\in\sum_{\beta<\gamma}\beta^2,$$

so $\alpha\in\gamma^3$, thus $\alpha\preccurlyeq\gamma^3$ for some $\gamma<\alpha$. However, this is clearly impossible since if $\gamma<\omega$, $\gamma^3\prec\omega\preccurlyeq\alpha$, otherwise

$$\gamma^3 = (\gamma^2)\gamma\approx\gamma\gamma\approx\gamma\prec\alpha.$$

This is a contradiction to $\alpha\prec\alpha^2$ and establishes our theorem.

COROLLARY. (a) *For every infinite ordinal α, $\alpha\approx\alpha+\alpha$.*

(b) *For every infinite ordinal α, $\alpha\approx\alpha+1$. Thus each infinite cardinal is a limit ordinal.*

Proof. If α is an infinite ordinal,

$$\alpha\preccurlyeq\alpha+1\preccurlyeq\alpha+\alpha\approx\alpha\times 2\preccurlyeq\alpha\times\alpha\approx\alpha^2\approx\alpha,$$

so $\alpha\approx\alpha+1\approx\alpha+\alpha$.

I wish to acknowledge the help of Professor M. Zuckerman and an unknown referee.

PROBLEMS AND SOLUTIONS

EDITED BY EMORY P. STARKE

All problems (both elementary and advanced) proposed for inclusion in this Department should be sent to E. P. Starke, 1000 Kensington Ave., Plainfield NJ 07060. Proposers of problems are urged to enclose any solutions or information that will assist the editors. Ordinarily, problems in well-known textbooks and results in generally accessible sources are not appropriate for this Department. No solutions (except those accompanying proposals) should be sent to Professor Starke.

ELEMENTARY PROBLEMS

Solutions of Elementary Problems should be sent to Problems Group, Mathematics Department, University of Maine, Orono, ME 04473. To facilitate their consideration, solutions of Elementary Problems in this issue should be typed (with double spacing) and should be mailed before January 31, 1971. Contributors (in the United States) who desire acknowledgment of receipt of their solutions are asked to enclose self-addressed stamped postcards.

E 2253. *Proposed by L. A. Gehami, Glastonbury, Connecticut*

Find the number S_n of distinct n-tuples $(a_1, a_2, \ldots, a_n)$ such that a_i is a nonnegative integer, $i=1, 2, \ldots, n$; $\sum_{i=1}^{n} a_i = n$; $\sum_{i=1}^{k} a_i \geqq k$, $k=1, 2, \ldots, n$.

E 2254. *Proposed by Marvin Marcus, University of California at Santa Barbara*

Let H be an $n\times n$ hermitian matrix and let $|H|$ be the matrix obtained from H by replacing each entry by its absolute value. Show that if $H\geqq 0$ (i.e., H is positive semi-definite) and $n\leqq 3$, then $|H|\geqq 0$. Show that for each $n\geqq 4$ there exists an $H\geqq 0$ such that $|H|$ is indefinite.

E2255. *Proposed by T. E. Mott, State University College, Buffalo, N. Y.*

Let $f(x_1, \cdots, x_n)$ be a real valued function defined on an open set $G\subseteq R^n$ and let $v_i=(\lambda_{i,1}, \cdots, \lambda_{i,n})$, $i=1, \cdots, n$, be linearly independent vectors in R^n. If the function f is continuous along that portion of every line passing through G and parallel to v_i, $i=1, \cdots, n$, and f is monotonic along each of these lines (the direction of monotonicity depending upon the choice of line), then $f(x_1, \cdots, x_n)$ is continuous in G.

E2256. *Proposed by H. Kestelman, University College, London, England*

Determine a real function f on R^n so that $|f(x)|\leqq\|x\|$ for all x where $\|x\|=(x_1^2+\cdots+x_n^2)^{1/2}$ and $\|\text{grad } f(0)\|=n^{1/2}$.

E 2257. *Proposed by I. Kaucký, Bratislava, Czechoslovakia*

Prove the identity

$$\sum_{k=1}^{2n-1}(-1)^{k-1}\binom{2n-1}{k}^{-1}\sum_{j=1}^{k}\frac{1}{j}=\frac{2n}{2n+1}\sum_{k=1}^{2n}\frac{1}{k}.$$

E 2258. *Proposed by Arthur Marshall, Madison, Wisconsin*

For each natural number k, let N_k be the kth number in the sequence which consists only of primes and products of consecutive primes, taken in natural order. Let the Möbius function be defined as usual: $\mu(1)=1$, and for $n>1$, $\mu(n)=(-1)^r$, where r is the number of primes dividing n. Prove

$$\sum_{k=1}^{\infty}\frac{\mu(N_k)}{N_k}=-\infty.$$

SOLUTIONS OF ELEMENTARY PROBLEMS

Route Determinism

E 1980 [1967, 438]. *Proposed by R. E. Chandler, Duke University*

The street plan for a certain town is a generalization of the tic-tac-toe board with n vertical streets and m horizontal streets (with no vertical street intersecting the end of a horizontal street and vice versa). An automobile will start at the end of one (unspecified) street and travel to the end of another. What is the minimum number of traffic counters (each to be placed in the middle of a block) necessary to determine the route of the automobile? The route is to have no cycles, and the counters record the time that the automobile passes so that the direction of the route is determined when two counters have been passed.

Solution by H. Shank, Cornell University. More generally, consider a connected graph G of v vertices and e edges to which $k \geqq 2$ additional "end edges" are attached, making a new graph G^*. To avoid special cases, we assume (a) G is inseparable (i.e., removing an edge from G never disconnects G) and (b) the end edges are attached at different vertices of G. (These conditions are satisfied in the problem.)

Now identify the k non-G end vertices with a single vertex n_0. This changes G^* into a new graph G_0^*.

We seek a minimal set S of branches such that (i) each loop (cycle) in G_0^* contains an edge of S, and (ii) each loop in G_0^* passing through n_0 contains at least two edges of S. (The edges in S contain the traffic counters; condition (ii) means the counters always determine direction of a trip across town.)

Let T be the complement of S, i.e., the set of branches of G_0^* not in S. By (i), T contains no loop. By standard graph theory, T is a forest of trees, so

$$|T| \leqq |\text{vertices}(G_0^*)| - 1 = (v+1) - 1 = v.$$

Hence

$$|S| \geqq |\operatorname{edges}(G_0^*)| - v = (e + k) - v = e - v + k.$$

Suppose $|S| = e - v + k$. Then T is a maximal tree of G_0^*, hence there exists a loop through n_0 containing *exactly one* edge not in T. This contradicts (ii), so we conclude that

$$|S| \geqq (e - v + 1) + k.$$

But there exists a set S which satisfies (i) and (ii) and has $|S| = (e - v + 1) + k$, namely, the set S consisting of the complement of any maximal tree in G and the k end edges. Hence

$$\min |S| = (e - v + 1) + k.$$

In the original problem, $v = mn$, $e = 2mn - m - n$, and $k = 2m + 2n$, so

$$\min |S| = mn + m + n + 1 = (m + 1)(n + 1).$$

Also solved by E. F. Schmeichel.

A more detailed solution will appear in Robert Kopp, Systems of roads with counters, this MONTHLY, probably December 1970.

The Greatest Integer as a Double Sum

E 2198 [1969, 1063]. *Proposed by Michael Aissen, Fordham University*

If $r > 1$ is an integer and x is real, define

$$f(x) = \sum_{k=0}^{\infty} \sum_{j=1}^{r-1} \left[\frac{x + jr^k}{r^{k+1}}\right],$$

where the brackets denote the greatest integer function. Show that

$$f(x) = \begin{cases} [x] & \text{if } x \geqq 0 \\ [x + 1], & \text{if } x < 0. \end{cases}$$

Solution by Howard Thoyre and George Kung, Wisconsin State University at Stevens Point. It is an easy exercise to show that if α is a real number and k is a positive integer,

$$\sum_{i=0}^{k-1} \left[\alpha + \frac{i}{k}\right] = [k\alpha].$$

Hence,

$$f(x) = \sum_{k=0}^{\infty} \sum_{j=1}^{r-1} \left[\frac{x}{r^{k+1}} + \frac{j}{r}\right] = \sum_{k=0}^{\infty} \left(\left[\frac{x}{r^k}\right] - \left[\frac{x}{r^{k+1}}\right]\right).$$

Letting S_n represent the nth partial sum, we have $S_n = [x] - [x/r^{n+1}]$. Since $r > 1$, there exists a positive integer N such that for all $n > N$, $|x/r^{n+1}| < 1$. Therefore, for fixed x and $n > N$, the sequence of partial sums is constant. For $x \geqq 0$ and $n > N$, the sequence is $[x]$ and for $x < 0$, the sequence is $[x] + 1 = [x + 1]$.

Also solved by M. T. Bird, D. M. Bloom, L. Carlitz, G. C. Dodds, D. F. Eliezer, M. G. Greening (Australia), Doug Hanto, Robert Heller, Dean Hickerson, Stephen Hoffman, James Inglis, Kenneth Klinger, R. S. R. Rao (India), Simeon Reich (Israel), Henry Ricardo, Bart Rice, Steve Rohde, E. F. Schmeichel, David Spear, T. Tamura (Japan), E. W. Trost (Switzerland), E. R. Van Eschen, Charles Wexler, and the proposer.

Integral Values for a Factorial Quotient

E 2199 [1969, 1063]. *Proposed by F. J. Papp, University of Delaware*

For given n, determine all r and s such that

$$\frac{(r+1)(n+s)!}{(s+1)(n+r)!}$$

is integral.

Solution by E. P. Del Norte, University of Texas at El Paso. For $n>1$ we have

$$R=\begin{cases}\dfrac{r+1}{(s+1)\cdot(n+s+1)(n+s+2)\cdots(n+r)} & s<r,\\ 1 & s=r,\\ \dfrac{(r+1)\cdots s\cdot(s+2)\cdots(s+n)}{(r+2)\cdots(r+n)} & r<s.\end{cases}$$

Thus R is integral for $s=r$ and $s\geqq r+n$. For $s<r$, R is not integral since $0<R<1$. Now let $r<s<r+n$ and set $s=r+k$:

$$R=\frac{(r+1)(r+n+1)(r+n+2)\cdots(r+n+k)}{r+k+1}.$$

But $r+k+1$ is a divisor of $(r+1)(r+n+1)\ldots(r+n+k)$ which equals

$$(r+k+1-k)(r+k+1+(n-k))\cdots(r+k+1+(n-1))$$

if and only if $r+k+1$ divides $k(n-k)(n-k+1)\ldots(n-1)$. It follows that for $r<s<r+n$ the pairs (r, s) for which R is integral are $(d-k-1, d-1)$ where $0<k<n$, and d divides $k(n-k)\ldots(n-1)$ and $d>k+1$.

For $n=1$, $R=s!/r!$ is integral whenever $s\geqq r$, and not integral for $s<r$.
For $n=0$, $R=(r+1)s!/(s+1)r!$. If $s<r$,

$$R=\frac{r+1}{(s+1)\cdot(s+1)(s+2)\cdots r},$$

which is not integral since $(r, r+1)=1$. $R=1$ for $s=r$. Now take $s>r$: $R=(r+1)^2(r+2)\cdots s/(s+1)$. Choose $k>0$ and let $s=r+k$. Proceeding as before we have that $r+k+1$ divides

$$\begin{aligned}&(r+1)^2(r+2)\cdots(r+k)\\ &\qquad=(r+k+1-k)^2(r+k+1+(1-k))\cdots(r+k+1-1)\end{aligned}$$

if and only if $r+k+1$ divides $k^2(k-1)!$, and hence R is integral for $(r, s) = (d-k-1, d-1)$, where $d \mid k^2(k-1)!$ and $d>k+1$.

Also solved by E. F. Schmeichel, and by Charles Wexler. Partial solution by Michael Goldberg.

Counting Unequal Partitions

E 2200 [1969, 1063]. *Proposed by J. M. Moser, Navy Electronics Laboratory, San Diego, Cal.*

Find the number of partitions of an integer n into m unequal parts such that each part must be not larger than $2m$ nor smaller than q, where $2 \leqq q \leqq m+1$.

Solution by E. F. Schmeichel, The College of Wooster. Set $q'=q-1$ and $k=m-q' \geqq 0$. Put $B=\binom{m+1}{2}$. Unless

$$\sum_{i=1}^{m} (q'+i) = mq' + B \leqq n \leqq m^2 + B = \sum_{i=1}^{m} (m+i),$$

there is no representation for n in the desired form. Otherwise, let $n=mq'+B+\alpha$, where $0 \leqq \alpha \leqq km$. It follows easily that the number of representations of n as a sum of m unequal parts chosen from $q'+1, q'+2, \cdots, q'+(m+k)=2m$ will be precisely the same as the number of representations of $B+\alpha$ as a sum of m unequal integers chosen from $1, 2, \ldots, m+k$. But

$$F_{m+k}(x, m) = x^B \frac{(1-x^{k+1})(1-x^{k+2}) \cdots (1-x^{k+m})}{(1-x)(1-x^2) \cdots (1-x^m)}$$

enumerates the partitions of n into m unequal parts, none of which exceed $m+k$. (See John Riordan, *Introduction to Combinatorial Analysis*, 1958, p. 153, Exercise 7.) So

$$\frac{F_{m+k}(x, m)}{x^B} = \frac{(1-x^{k+1})(1-x^{k+2}) \cdots (1-x^{k+m})}{(1-x)(1-x^2) \cdots (1-x^m)} = \sum_{\alpha=0}^{km} P_\alpha x^\alpha,$$

where P_α denotes the number of partitions of $n=mq'+B+\alpha$ into m unequal parts ranging between q and $2m$.

Also solved by Michael Goldberg, C. T. Haskell, and W. F. Rogers. Letting $p'(n, m; q, k)$ denote the number of partitions of n into m unequal parts, each part $\geqq q$ and $\leqq k$, and letting $v(n, m; r)$ denote the number of partitions of n into at most m parts, each part $\leqq r$. Haskell first proves that $p'(n, m; q, k) = v(n-mq-m(m-1)/2, m; k-m-q+1)$.

Steiner-Artzt Coincidence

E 2201 [1969, 1063]. *Proposed by J. M. Quoniam, Saint-Etienne, France*

Given a triangle, find (with compass and straightedge) the points in which the inscribed Steiner ellipse of the triangle intersects the three Artzt parabolas of the triangle, and also construct the tangents to the curves at these points. (The *inscribed Steiner ellipse* of a triangle is the ellipse inscribed in the triangle and

having the centroid of the triangle for its center; an *Artzt parabola* is the parabola tangent to two sides of the triangle at the endpoints of the third side.)

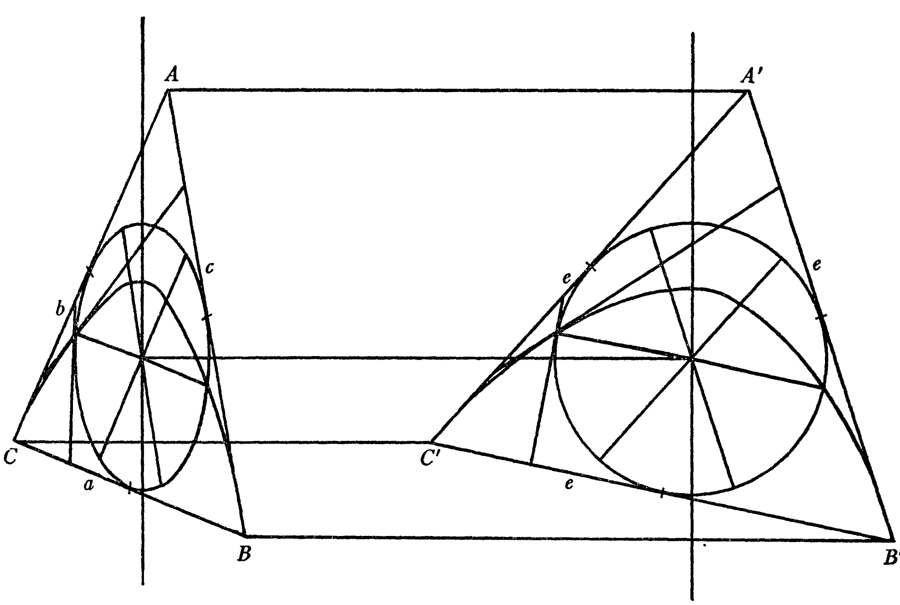
I. *Solution by A. W. Walker, Toronto, Canada.* The two points where the Artzt (A) parabola meets the inscribed Steiner ellipse are distant $(BC)/(2\sqrt{3})$ from the centroid G of the triangle ABC on a line through G parallel to BC; the tangents to the parabola at these points meet at the midpoint of AG, and the tangents to the ellipse are parallel to AG. These results are readily verified when ABC is equilateral, the parabola then having G as focus and the right bisector of AG as directrix, and follow generally by affine projection.

II. *Solution by Michael Goldberg, Washington, D.C.* By orthogonal projection, project the given triangle ABC into an equilateral triangle. If the lengths of the sides of the given triangle are a, b, c, where $a \leqq b \leqq c$, then the length of the side of the equilateral triangle is e, where e satisfies the equation

$$\sqrt{e^2 - a^2} = \sqrt{e^2 - b^2} + \sqrt{e^2 - c^2},$$

or $3e^4 - 2(a^2+b^2+c^2)e^2 + 16K^2 = 0$, where K is the area of the given triangle. Since e is obtainable from a, b, c by the taking of square roots, it is constructible by Euclidean methods.

For the equilateral triangle, the Steiner ellipse is the inscribed circle. It is easily verified that an Artzt parabola intersects this circle at the ends of a diameter that is parallel to a side. The tangent to the circle at such a point is perpendicular to a side. The tangent to the parabola makes an angle of 45° with a side.



If the equilateral triangle is projected back into the given triangle then the circle is projected into the Steiner ellipse of the given triangle, and the Artzt parabolas are projected into the Artzt parabolas of the given triangle. The tangents are projected into the desired tangents in the given triangle. The determination of all the desired points and lines requires only Euclidean constructions.

Also solved by James Buddenhagen, Charles Chouteau, Jordi Dou (Spain), M. G. Greening (Australia), and the proposer.

The Transition Span of a Set of Ordered Pairs of Reals

E 2202 [1969, 1137]. *Proposed by Necdet Üçoluk, Clarion (Pa.) State College*

Since a binary relation on a set A is a subset of $A\times A$, every subset of the Euclidean plane E_2 is a binary relation on the reals. Define the *transitive span* of a relation ρ to be the smallest transitive relation containing ρ. (a) Determine the transitive span of the relation ρ, consisting of all points interior to or on the triangle with vertices (3, 5), (2, 7) and (6, 0). (b) What if the region is the set bounded by the ellipse $(x-1)^2+4y^2=16$?

Solution (a) by Grattan Murphy, University of Maine. Let A be a set and S a relation in A. For any sets P, $Q\subset A\times A$, let $P\circ Q=\{(x, y):(x, a)\in P$ and $(a, y)\in Q$ for some $a\in A\}$. Then define $S^1=S$ and $S^n=S^{n-1}\circ S$ for each $n>1$. It is easy to see that $(x, y)\in S^n$ if and only if there exist $a_1, \cdots, a_{n-1}\in A$ such that (x, a_1), (a_1, a_2), $\cdots$, $(a_{n-1}, y)\in S$, from which $S^p\circ S^q=S^{p+q}$. Now any transitive relation containing S must contain S^n for every n and, therefore, also $\bigcup S^n$. But $\bigcup S^n$ is transitive, for if (x, y) and $(y, z)\in\bigcup S^n$, then there exist p, q such that $(x, y)\in S^p$ and $(y, z)\in S^q$, whence $(x, z)\in S^p\circ S^q=S^{p+q}\subset\bigcup S^n$. Therefore $\bigcup S^n$ is the transitive span of S.

Let S be a closed convex set in E^2 such that the lines $y=a$ and $x=a$ intersect S in intervals bounded by the points (x_1, a), (x_2, a), (a, y_1) and (a, y_2) with $x_1<x_2$, $y_1<y_2$. Then the rectangular region $\{(x, y)\,|\,x_1\leqq x\leqq x_2,\ y_1\leqq y\leqq y_2\}$ is in $S\circ S$. Let the boundary at (x_1, a) and (a, y_1) satisfy $x=f_1(y)$ and $x=f_2(y)$, respectively. Then $f_2(y_1)=a$ and $x_1=f_1(a)=f_1(f_2(y_1))$. The lower left vertex of the rectangle is $(f_1\circ f_2(y_1), y_1)$. The boundary points of $S\circ S$ must be vertices of a rectangle of this form. Similarly the boundary points of S^n are of the form $(f_1\circ f_2\circ\cdots\circ f_n(y), y)$ where $x=f_i(y)$ are appropriately chosen boundary curves of S.

In the case where S is the triangle with vertices (2, 7), (3, 5), and (6, 0), the boundary curves are $f(y)=x=6-4y/7$, $g(y)=x=(11-y)/2$, and $h(y)=x=6-3y/5$. The right and left hand boundaries of S^{2n} are $x=(fh)^n(y)$, $x=(fh)^{n-1}fg(y)$, $x=(hf)^{n-1}gf(y)$ and $x=(hf)^n(y)$, while those of S^{2n+1} are $x=f(hf)^{n-1}gf(y)$, $x=f(hf)^n(y)$, $x=(hf)^nh(y)$ and $x=(hf)^ng(y)$ for $n=1, 2, \cdots$; the upper and lower boundaries being $y=7$ and $y=0$ in each case. The slopes of the boundary lines of S^{2n} are all positive and tend to ∞ as $n\to\infty$ with the points (90/23, 90/23) and (84/23, 84/23) being fixed points for right and left boundaries, respectively. Likewise the slopes of the boundaries of S^{2n+1} tend to $-\infty$ and contain the points

(90/23, 84/23) and (84/23, 90/23). The points on the open intervals on the lines $x=84/23$ and $x=90/23$ between these points are not contained in any S^n.

Solution (b) by E. F. Schmeichel, College of Wooster, Ohio. Let E denote the ellipse and its interior. Through the point $(1, 1)\in E$ draw horizontal and vertical lines intersecting E in segments PP' and QQ', respectively, and let $ABCD$ be the rectangle having Q and Q' on its horizontal sides and P and P' on its vertical sides. Note that rectangle $ABCD$ must be in the transitive span of E. Repeat the above process with the point $(0, 0)\in E$ and the set $ABCD\cup E$ to obtain rectangle $A'B'C'D'$, a transitive set and the desired transitive span of E. (Here $A'=(-3, 2)$, $B'=(5, 2)$, $C'=(5, -2)$, and $D'=(-3, -2)$.)

Also solved (part (b)) by J. P. Comiskey.

Consequences of a Lucas Congruence

E 2205 [1969, 1138]. *Proposed by S. M. Farber, D. W. Walkup, and R. J. B. Wets, Boeing Scientific Research Laboratories*

Suppose nonnegative integers m and n are given in their representations to a prime base p, i.e.

$$m = (r_k \cdots r_1 r_0)_p = \sum_{i=0}^{k} r_i p^i, \qquad 0 \leqq r_i < p,$$

$$n = (s_k \cdots s_1 s_0)_p = \sum_{i=0}^{k} s_i p^i, \qquad 0 \leqq s_i < p.$$

Find a simple expression for the binomial coefficient $C(m, n)=\binom{m}{n}$ mod p. In particular, find necessary and sufficient conditions for

$$\text{(a)}\quad \binom{m}{n} \equiv 0 \mod p \qquad \text{(b)}\quad \binom{2m-1}{m-1} \text{ odd.}$$

Solution by Robert Fray, Florida State University. A well-known result due to E. Lucas says that $C(m, n)\equiv C(r_0, s_0)C(r_1, s_1)C(r_2, s_2) \cdots C(r_k, s_k) \pmod p$. See L. E. Dickson, *History of the Theory of Numbers*, 1919, vol. 1, p. 271; also N. J. Fine, *Binomial coefficients modulo a prime*, this MONTHLY, 54(1947), pp. 589–592.

(a) Since $0\leqq r_i, s_i<p$, the binomial coefficient $C(r_i, s_i)$ is divisible by p if and only if $s_i>r_i$. Hence $C(m, n)\equiv 0 \pmod p$ if and only if $s_i>r_i$ for some i, where $0\leqq i\leqq k$.

(b) If we choose $p=2$, and let $m-1=\sum_{i=0}^{k} r_i 2^i$, where r_i is 0 or 1 for $0\leqq i<k$, and $r_k=1$, then $2m-1=1+\sum_{i=1}^{k+1} r_{i-1}2^i$. It follows that

$$C(2m-1, m-1) \equiv C(1, r_0)C(r_0, r_1)C(r_1, r_2) \cdots C(r_{k-1}, r_k)C(r_k, 0) \pmod 2.$$

Hence $C(2m-1, m-1)$ is odd if and only if $1\geqq r_0\geqq r_1\geqq r_2\geqq \cdots \geqq r_{k-1}\geqq r_k=1$,

i.e. $m-1=1+2+2^2+2^3+\cdots+2^k=2^{k+1}-1$. Since $C(1, 0)$ is odd, it follows that $C(2m-1, m-1)$ is odd if and only if m is a nonnegative power of 2.

The Lucas congruence has been extended to the q-binomial coefficient. The result in part (a) of this problem has been proved for the multinomial coefficient, the q-binomial coefficient, and the q-multinomial coefficient. (See R. D. Fray, *Congruence properties of ordinary and q-binomial coefficients*, Duke Math. J., vol. 34(1967), pp. 467–480.)

Also solved by L. Carlitz, N. J. Fine, Michael Goldberg, Wells Johnson, Douglas Lind, D.C.B. Marsh, Simeon Reich (Israel), E. F. Schmeichel, Michael Shimshoni (Israel), Ira Sterbakov, T. Tamura (Japan), Charles Wexler, and the proposers.

ADVANCED PROBLEMS

All solutions of Advanced Problems should be sent to J. Barlaz, Rutgers—The State University, New Brunswick, N.J. 08903. Solutions of Advanced Problems in this issue should be typed (with double spacing) on separate, signed sheets and should be mailed before January 31, 1971. Contributors (in the United States) who desire acknowledgment of receipt of their solutions are asked to enclose self-addressed stamped postcards.

An asterisk (*) means neither the proposer nor the editors supplied a solution.

5753.* *Proposed by V. A. McAuley, Huntsville, Alabama*

Show that for all integers $n>2$,

$$\bar{x} = \left[\frac{n(22n-19)}{(n-2)(22n-31)}\right]^{1/2}$$

is an approximation to the positive real root of the equation

$$(n-2)x^n - nx^{n-2} - 2 = 0,$$

with at most an error in the third decimal place.

5754. *Proposed by M. S. Klamkin, Ford Scientific Laboratory*

If $L(a, b)$ denotes the perimeter of an ellipse with semi-axes a and b ($a \geqq b$), show that

$$L^2(a, c) - 16a^2 = L^2(b, c) - 16b^2.$$

5755. *Proposed by G. G. Kiziak, University of Toronto*

It is well known that if a sequence of commuting normal operators T_n on a Hilbert space H converges strongly to an operator T, then T is normal. Show that the assumption that the operators commute cannot be dropped, even if the T_n are unitary. (This contradicts an assertion in Nagy and Foias, *Analyse Harmonique des Operateurs de l'espace de Hilbert*, p. 107, paragraph 2.)

5756.* *Proposed by J. H. Westbrooke, Evanston, Ill.*

Let $q_1=13$, $q_2=17$, and $q_3, q_4, \cdots$ be the successive primes of the form $4k+1$. Find the least upper bound of the sequence whose nth term is

$$1-\prod_{k=1}^{n}\left(1-\frac{2}{q_k}\right).$$

5757. *Proposed by R. A. Struble, North Carolina State University*

Does there exist a solution of the boundary value problem:

$$\frac{dy}{dx}+\sin y=\int_{t/2}^{t}[1+y^2(s)]\sin s\,ds, \qquad y(0)=y(1)?$$

If so, how many?

5758. *Proposed by W. O. Egerland, Research and Development Center, Aberdeen Proving Grounds, Md.*

A theorem of E. Landau states: If $f(z)$ is analytic and bounded by 1 in the unit disk, $|z|<1$, and satisfies the conditions $f(0)=0$ and $f'(0)=a$, $0<a<1$, then $f(z)$ is univalent in the disk $|z|<\rho=a^{-1}(1-(1-a^2)^{1/2})$. What is the corresponding ρ if, in addition, $f(z)\neq 0$ for $z\neq 0$?

SOLUTIONS OF ADVANCED PROBLEMS

Ideals in an Associative Ring

5675 [1969, 565]. *Proposed by M. Slater, University of Bristol, England*

Let R be an associative ring, A an ideal of R, and U a nonzero ideal of A. Under suitable conditions U contains a nonzero ideal of R (cf. Jacobson, *Structure of Rings*, p. 65). Construct an example in which the conclusion is false, and in addition (1) A is as well-behaved as possible; (2) $U^2\neq(0)$; (3) $U^2\neq(0)$ and $A^2=A$.

Solution by the proposer. All our examples will be algebras over any field F. We write $\langle x_1, x_2, \cdots, x_n\rangle$ for the linear span of the indicated elements, and use of the notation indicates that these elements are linearly independent.

(1) AUA is an ideal of R contained in U; hence is (0). If A has no total right divisors of zero, then $UA=(0)$, and $U\neq(0)$ consists of left total divisors of zero in A. Thus, in any example A must have (say) left total divisors of zero. We give an example in which A has a left unity e; then A has the additional regularities $A^2=A$ and $Aa=(0)$, where $a\in A$, implies $a=0$.

$R=\langle e, a, u, s\rangle$; $A=\langle e, a, u\rangle$; $U=\langle u\rangle$. Basis products zero except $ee=e$, $ea=a$, $eu=u$, $us=a$.

(2) We assume R is an algebra over F. Then it can be shown that $\dim U\geqq 4$. If $\dim U=4$, say $U=\langle v, z, u, w\rangle$. Then it can be shown that $\dim A\geqq 7$; say $A\supseteq\langle a, b, c, v, z, u, w\rangle$, and of course $\dim R\geqq 8$; say $s\in R$, $s\notin A$. By suitable

choices of the indicated elements we can present a partial multiplication table for R in the following form:

	a	b	c	d	v	z	u	w	s	p	t
a	Y	Y	0	Y	Y	Y	0	0			
b	$c+Y$		0		$z+Y$	X	0	0			
c	0	0	0	0	0	0	0	0			
d	Q		0		X	X	0	0			
v	u	Y	0	Y	w	0	0	0	a		
z	0	0	0	0	0	0	0	0	c		
u	0	0	0	0	0	0	0	0	Y	Y	Y
w	0	0	0	0	0	0	0	0	u	Y	Y
s							Q	X			
p							$c+Y$	$z+Y$			
t								X			

$Q=\langle c, u, w\rangle$; $X=\langle z, u, w\rangle$; $Y=\langle u, w\rangle$.

Capital letters in the table denote (unknown) elements in the corresponding subspace; p need not be independent of $\{s\}\cup A$; d and t are arbitrary elements of A and R respectively.

There actually exist examples of this type; two are as follows:

(i) $$R = \langle a, b, c, v, z, u, w, s\rangle; \ A \text{ and } B \text{ as above.}$$

Nonzero basis products: $ba=c$, $bv=z$, $va=u$, $vv=w$, $vs=a$, $zs=c$, $ws=u$, $sv=b$, $su=c$, $sw=z$.

(ii) $$R = \langle a, b, c, v, z, u, w, s, p\rangle; \ A \text{ and } B \text{ as in } (i).$$

Basis products as in (i) except that the last three listed are replaced by $pu=c$, $pw=z$.

(3) We do not have a solution for (3), but we can prove that in an example with $A^2=A$ we must have dim $U>4$. (In the examples of (2) we have $A^2=(0)$.) If dim $U\leqq 4$, then R has a sub-multiplication table as shown in (2). Suppose $b\in A^2$. Then $b=\sum f_i g_i$, say, with f_i, $g_i\in A$. Then $c+Y=ba=(\sum f_i g_i)a = \sum f_i(g_i a) = \sum f_i q_i=0$ from the table ($q_i\in Q$). Since $c\notin Y$, this is a contradiction. So, if dim $U\leqq 4$, $b\notin A^2$, and so $A^2\neq A$.

Nonlinear Homogeneous Isometries

5688 [1969, 835]. *Proposed by P. R. Chernoff, University of California, Berkeley*

Let X and Y be normed vector spaces and $f: X\to Y$ a homogeneous isometry; that is: $f(t\cdot v)=tf(v)$ and $\|f(v)-f(w)\|=\|v-w\|$ for all scalars t and vectors v, w. Must f be linear? (Cf. Banach, p. 166, Theorem 2.)

Solution by J. Marsden, University of California, Berkeley. If Y is two dimensional or strictly convex or if f is onto, then f is linear; but in general f need not be linear.

We first give the example. Let $X=\mathbf{R}^2$ and $Y=\mathbf{R}^3$, each with the L_∞ norm (equalling the supremum of the coordinate norms). Consider the piecewise linear homogeneous function $f: X\to Y$ defined by $(\pm 1, 0)\to(\pm 1, 0, 0)$, $(0, \pm 1)\to(0, \pm 1, 0)$ and $(1, \pm 1)\to(1, \pm 1, \pm 1)$. It is straightforward to check that f is an isometry.

That f is linear if it is onto is the cited theorem of Banach. To show f is linear if Y is two dimensional, one is quickly reduced to the case in which X is two dimensional. But by invariance of domain, f is open so, by homogeneity, is onto, and therefore linear.

Finally, if Y is strictly convex, (i.e. equality in the triangle inequality holds only for collinear points), the equalities

$$\|f(u)-f(v)\| = \left\|f(u)-f\left(\frac{u+v}{2}\right)\right\| + \left\|f(v)-f\left(\frac{u+v}{2}\right)\right\|,$$

$$\left\|f(u)-f\left(\frac{u+v}{2}\right)\right\| = \left\|f(v)-f\left(\frac{u+v}{2}\right)\right\|$$

imply

$$f\left(\frac{u+v}{2}\right) = \frac{f(u)+f(v)}{2}$$

from which linearity is evident.

Also solved by P. J. Owens (England), T. V. Wimer, and the proposer.

Zero Divisors in an Associative Ring

5690 [1969, 947]. *Proposed by Irving Kaplansky, University of Chicago*

Let R be an associative ring. Suppose $a\in R$ is nilpotent and $b\in R$ is a right zero-divisor. Suppose a and b commute. Prove there exists $c\neq 0$ in R such that $ac=bc=0$.

Solution by Barbara Osofsky, Rutgers—The State University. Assume $d\neq 0$ such that $bd=0$. Then, since there exists n such that $a^n=0$, there exists $k\geqq 0$ with $a^kd\neq 0$, $a^{k+1}d=0$. Now $a\cdot a^kd=0$ and $ba^kd=a^kbd=0$: that is, $c=a^kd$ establishes the result.

If we interpret "b is a right zero-divisor" to mean: there exists $d\neq 0$ such that $db=0$, then the statement of the problem would be false.

Let $(R, +)$ be the additive group of $S[z]$, where $S=F[x, y]/(x^2)$, F some field. That is, $R=\left\{\sum_{i=0}^n \alpha_i(x, y)z^i\right\}$, where α_i is an at most linear polynomial in x and any degree in y. Define multiplication in R by the distributive law and the $\alpha(x, y)$ polynomials multiply as in S, $z\alpha=\alpha z$ for $\alpha\in F$, $zx=xz$, and $zy=0$.

Now x and y commute, $x^2=0$, and $zy=0$, $z\neq 0$, yet for all nonzero e in R, $ye\neq 0$.

Also solved by G. A. Elliott, Harry Gonshor, Melvin Hendriksen, D. A. Higgs, M. L. Laplaza (Puerto Rico), Bruce Jensen & Ernest Sparks, C. O. M. Pewter (England), Peter Ross, G. F. Schumm, A. J. Silberger, James H. Smith, Seth Warner, A. K. Wayman, Jr., Qazi Zameeruddin (India), and the proposer.

A Discrete Set with Uncountable Closure

5691 [1969, 947]. *Proposed by John Bond and Harold Reiter, Clemson University*

Does there exist a discrete subset of the real line (usual topology) which has an uncountable closure?

Solution by J. C. Morgan II, University of California, Berkeley. The first example of such a set was given implicitly by I. Bendixon (*Quelques Théorèmes de la Théorie des Ensembles de Points*, Acta Math., 2(1883) pp. 416–418). Namely, consider any nowhere dense (nonempty) perfect set K. In each complementary interval of K choose a decreasing sequence converging to the left end-point and an increasing sequence converging to the right endpoint. The set of all sequences chosen is a discrete set whose closure contains K.

Also solved by sixty other readers.

Editorial Note. Many contributors pointed out that it suffices to select the midpoints of the complementary intervals of Cantor's ternary set. Ter Morsche and van der Steen note—as implied above—that the closure may also have positive measure.

Characteristic Roots of a Symmetric Matrix

5692 [1969, 948]. *Proposed by Olga Taussky, California Institute of Technology*

Let F be a real closed field and A a finite matrix with elements in F. Show explicitly that the characteristic roots of $A'A$ (A' the transposed matrix) are sums of squares in F.

Solution by J. B. Skinner, University of Massachusetts. Since F is real closed, the extension $F(\sqrt{-1})$ is proper and algebraically closed. There is an F-automorphism of $F(\sqrt{-1})$ given by $z=x+y\sqrt{-1}\to\bar{z}=x-y\sqrt{-1}$, and the space of n-tuples over $F(\sqrt{-1})$ has an inner product given by $(x, y)=\sum_{i=1}^{n} x_i\bar{y}_i$. If λ is a characteristic root of $A'A$, then $\lambda\in F(\sqrt{-1})$; furthermore, since $A'A$ is symmetric, it follows that $\lambda\in F$, hence $A'A$ has a characteristic vector with coordinates in F. We may assume that $(x, x)=1$. Then $\lambda=\lambda(x, x)=(x, \lambda x)=(x, A'Ax)$. But this last expression equals (Ax, Ax), which is a sum of squares in F.

Also solved by G. N. de Oliveira (Portugal), E. D. Dixon, R. A. Kopas, C. L. Sabharwal, R. C. Thompson, E. W. Trost (Switzerland), and the proposer.

The proposer proves also that the characteristic roots are sums of squares in a formally real field if we assume that F is a formally real field.

Metric in a Half-open Interval Topology

5694 [1969, 948]. *Proposed by Jürg Rätz and Albert Wilansky, Lehigh University*

As a subspace of R (the reals) with the half-open interval topology, (Kelley, *General Topology*, p. 59), Q (the rationals) is first countable, hence second countable, hence metrizable. Exhibit an explicit metric for this topology.

Solution by D. R. Anderson, Central Washington State College. Choose a listing of the rationals, $q_1, \cdots, q_n, \cdots$. For each positive integer n, the function

$$d_n(x, y) = \begin{cases} |x - y| & \text{if } x, y < q_n \quad \text{or} \quad x, y \geqq q_n \\ |x - y| + 1 & \text{if } x < q_n \leqq y \quad \text{or} \quad y < q_n \leqq x \end{cases}$$

is a metric for R for which the open sets are the usual open sets and the half open intervals of the form $[q_n, b)$. The function

$$d(x, y) = \sum_{n \geqq 1} d_n(x, y)/n^2$$

is also seen to be a metric for R for which the open sets include the usual open sets and the half open intervals of the form $[q_n, b)$, $n=1, 2, \cdots$. In fact, these are the only open sets. This metric may then be restricted to the rationals in order to obtain the desired metric.

Also solved by Linda W. Brinn, E. P. Del Norte, R. J. Driscoll, G. J. Foschini, T. E. Gantner. Leonard Gillman, Gary Gruenhage, J. B. Lindner, O. P. Lossers (Netherlands), M. D. Mavinkurve (India), Ka Menehune, D. Hammond Smith (England), R. H. Sorgenfrey, T. Wimer, and the proposer.

Boolean Lattices

5696 [1969, 1074]. *Proposed by P. J. Chase, Laurel, Maryland*

Show that a lattice L is Boolean if and only if it admits a unary operation $x \to x'$ such that $a \cap b \leqq c \leqq a \cup b$ implies $c \cap b' \leqq a \leqq c \cup b'$.

Solution by Michael McCoy, Kansas State Teachers College. First suppose that L is a Boolean lattice, and let complementation be the operation $x \to x'$ (well-defined since complements are unique in a Boolean lattics.) Using the distributive property and the isotonicity of the meet and join operations: if $a \cap b \leqq c \leqq a \cup b$, then

$$\begin{aligned} c \cap b' &\leqq a \cup (c \cap b') = (a \cup c) \cap (a \cup b') \leqq (a \cup b) \cap (a \cup b') \\ &= a \cup (b \cap b') = a \cup 0 = a = a \cap I = a \cap (b \cup b') \\ &= (a \cap b) \cup (a \cap b') \leqq (a \cap c) \cup (a \cap b') \\ &= a \cap (c \cup b') \leqq c \cup b'. \end{aligned}$$

Thus the necessity of the condition follows.

To prove the sufficiency suppose that $a\cap b\leqq c\leqq a\cup b$ implies $c\cap b'\leqq a\leqq c\cup b'$. If $b\in L$, then $b\cap a\leqq b\leqq b\cup a$ for all $a\in L$, and this implies $b\cap b'\leqq a\leqq b\cup b'$ for all $a\in L$; hence $b\cap b'=0$ and $b\cup b'=I$, which shows that L is complemented. Moreover, if a and b are complementary, then $a\cap b\leqq c\leqq a\cup b$ for all $c\in L$, which implies $b'\cap c\leqq a\leqq b'\cup c$ for all $c\in L$. But if $c=I$, then $b'=b'\cap I\leqq a$; and if $c=0$, then $a\leqq b'\cup 0=b'$. Therefore $a=b'$, and L is uniquely complemented. To show distributivity, we use the well-known characterization: a lattice is distributive if and only if relative complements are unique. Now $x\cap a=y\cap a$ and $x\cup a=y\cup a$ implies $y\cap a\leqq x\leqq y\cup a$ and $x\cap a\leqq y\leqq x\cup a$. By the hypothesis, $y\cap a\leqq x\leqq y\cup a$ implies $a'\cap x\leqq y\leqq a'\cup x$, which implies $x'\cap y\leqq a'\leqq x'\cup y$; and $x\cap a\leqq y\leqq x\cup a$ implies $x'\cap y\leqq a\leqq x'\cup y$. Together, these imply $x'\cap y=0$ and $x'\cup y=I$ which implies $x=y$ since L is uniquely complemented. Therefore L is complemented and distributive; hence, Boolean.

Also solved by T. S. Blyth (Scotland), N. A. Borba, D. Ž. Djokoviç, G. A. Edgar, S. I. Gendler, Khee-meng Koh, Yataro Matsushima (Japan), V. V. Rama Rao (India), Simeon Reich (Israel), and the proposer.

Dense Subsets in a Separable Space

5698 [1969, 1074]. *Proposed by D. L. Lutzer, University of Washington*

It is well known that a subspace S of a separable Hausdorff space X need not be separable. Is there an example of this in which S is dense?

Solution by Dennis Henkel, Allen-Bradley Company, Milwaukee, Wisconsin. Let X be the real numbers, and define a new topology so that a set V is open if it is empty or if it is of the form $V=U-K$ where U is open in the usual sense and K is a set of irrational numbers that is at most countable. The rational numbers are dense in this space and the set S of irrational numbers is also dense. However, the complement of any countable subset of S is open and nonempty in the relative topology.

Also solved by Harold Bennett, W. E. Bonnice (Turkey), Helen F. Cullen, T. E. Gantner, Leonard Gillman, K. D. Joshi, R. H. Lohman, Dan Marcus, T. M. Phillips, Jürg Rätz & Albert Wilansky, George Reynolds, D. H. Smith (England), Gerald Wildeberg, T. V. Wimer, and the proposer.

Derivation on a Ring of Continuous Functions

5699 [1969, 1074]. *Proposed by G. J. Foschini, Bell Telephone Laboratories.*

A derivation on a ring S is an additive mapping $s\to s'$ of S into itself satisfying $(pq)'=pq'+p'q$. Let $C[0, 1]$ be the ring of continuous real functions on $[0, 1]$ with the usual norm. Show that there exists a ring $R[0, 1]\subset C[0, 1]$ with a derivation $r\to r^{[1]}$ such that the following subsets of $R[0, 1]$ are dense in $C[0, 1]$:

(i) $\{r \mid r^{[1]} = 0\}$,

(ii) $\{r \mid r^{[1]} = r\}$,

(iii) $\{r \mid r^{[n]} = (r^{[n-1]})^{[1]} > 0,\ n = 1, 2, \cdots\}$

and such that $R[0, 1] \supset C^\infty[0, 1]$ whereon the derivation coincides with the usual derivative.

Solution by the proposer. Let $f \in C[0, 1]$ which is differentiable almost everywhere in such a way that there exists a $g \in C[0, 1]$ such that $g = df/dt$ almost everywhere. In such a case we define $g = f^{[1]}$. In the ring

$$R[0, 1] = \{r \in C[0, 1] \mid r^{[n]} \text{ exists}, n = 1, 2, \cdots\}$$

$r \to r^{[1]}$ is a derivation. This follows from the continuity of r and $r^{[1]}$ and the fact that $r^{[1]} = dr/dt$ almost everywhere.

To verify (i) we note that these constants are precisely the algebra of continuous singular functions. Let $C(t)$ denote the Cantor function. Given any pair (x, y) of distinct points, a clock speed λ, $0 < \lambda < 1$, can be chosen so that $C(\lambda t)$ separates x and y. Thus by the Stone-Weierstrass theorem, the constants are dense in $C[0, 1]$.

Let h denote an arbitrary element in $C[0, 1]$. By (i) we can choose sequences of constants $\{a_m\}$ and $\{b_m\}$ converging to e^{-t} and h respectively. Thus $a_m b_m e^t \to h$. Now $a_m b_m$ is constant and so $(a_m b_m e^t)^{[1]} = a_m b_m e^t$ is a fixed point and (ii) has been demonstrated.

Finally $(e^{t/m} - 1) + b_m \to h$ and $[(e^{t/m} - 1) + b_m]^{[n]} = e^{t/m}/m^n > 0$, $n = 1, 2, \cdots$. Therefore (iii) is satisfied.

Also solved by O. P. Lossers (Netherlands).

REVIEWS

EDITED BY KENNETH O. MAY

COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER

Beginning with the January 1971 issue this section will be edited by J. Arthur Seebach, Jr. and Lynn A. Steen with the collaboration of the Mathematics Departments at St. Olaf College and Carleton College.

All material for review should be sent to: Book Review Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

All unsigned material is written by the editors. A boldface C in the margin indicates tha a review is based in part on classroom use.

What Is Modern Mathematics. By G. Choquet. Educational Explorers Limited, Reading, England, 1963. x+46 pp. $2.00. (Telegraphic Review, April 1970.)

This is a reverential description of the Bourbaki attitude to mathematics, slightly colored by the author's personal preferences.

THE AMERICAN MATHEMATICAL MONTHLY

(FOUNDED IN 1894 BY BENJAMIN F. FINKEL)

THE OFFICIAL JOURNAL OF

THE MATHEMATICAL ASSOCIATION OF AMERICA, INC.

VOLUME 77 NUMBER 9

CONTENTS

(Continued on inside cover)

NOVEMBER 1970

NOTICE TO AUTHORS

Articles should be typewritten on good quality paper, triple spaced with wide margins. Submit the original and a duplicate; keep a complete copy for protection against loss. Please follow the format used in current issues of the Monthly. The manuscript must make the author's intent clear to the printer. A paper accepted for publication by the editor but not acceptable to the printer will be returned to the author for retyping. Please use the American Mathematical Society *Manual for Authors of Mathematical Papers* as your guide and please conform to correct English usage. A *Manual for Monthly Authors* is available on request.

Prospective Monthly authors are advised to consult the Statement of Policy in the January 1969 issue, p. 2. Specialized research is usually unsuitable. Backlog: Main Articles 9 months, Math. Notes 9 months, Research Problems 7 months, Classroom Notes 7 months, Math. Education 8 months.

EDITORIAL CORRESPONDENCE AND MAIN ARTICLES: to Harley Flanders, American Mathematical Monthly, Tel Aviv University, Ramat Aviv, Israel (see Notice, p. 555); NOTES, etc.: to the corresponding Associate Editor; ADVERTISING CORRESPONDENCE: to Raoul Hailpern, Mathematical Association of America, SUNY at Buffalo, Buffalo, N. Y. 14214; CHANGE OF ADDRESS and SUBSCRIPTIONS: to A. B. Willcox, Mathematical Association of America, 1225 Connecticut Ave., N.W., Washington, D. C. 20036.

Annual dues for members of the Association (including a subscription to the American Mathematical Monthly) are $10.00. For nonmembers the subscription price is $15.00.

Published by the Association at Washington, D. C., and Menasha, Wisconsin during the months of January February, March, April, May, June–July, August–September, October, November, December.

Second-class postage paid at Washington, D. C., and additional mailing offices. Acceptance for mailing at special rate of postage provided for in the Act of February 28, 1925, embodied in Paragraph 4, Section 538, P. L. and R., authorized April 1, 1926.

PRINTED IN THE UNITED STATES OF AMERICA

PITFALLS IN COMPUTATION, OR WHY A MATH BOOK ISN'T ENOUGH

GEORGE E. FORSYTHE, Computer Science Dept., Stanford University

1. Introduction. Why does a student take mathematics in college or university? I see two reasons: (i) To learn the structure of mathematics itself, because he (or she) finds it interesting. (ii) To prepare to apply mathematics to the solution of problems he expects to encounter in his own field, whether it be engineering, physics, economics, or whatever.

Surely (ii) motivates far more students than (i). Moreover, most solutions of major mathematical problems involve the use of automatic digital computers. Hence we may justifiably ask what mathematics courses have to say about carrying out mathematical work on a computer. This question motivates my paper.

I am not in a mathematics department, and sometimes I moralize about them. If the reader prefers not to be lectured to, let him ignore the preaching and just pay attention to the numerical phenomena for their own sake.

I want to acknowledge the help of Mr. Michael Malcolm in criticizing the manuscript and doing the computations with a special floating-decimal arithmetic simulator he wrote for Stanford's hexa-decimal computer, an IBM 360/67.

2. Nature of computers. An automatic digital computer is a general-purpose machine. The bits of information in its store can be used to represent any quantifiable objects—e.g., musical notes, letters of the alphabet, elements of a finite field, integers, rational numbers, parts of a graph, etc. Thus such a machine is a general abstract tool, and this generality makes computer science important, just as mathematics and natural language are important.

In the use of computers to represent letters of the alphabet, elements of a finite field, integers, etc., there need be no error in the representation, nor in the processes that operate upon the quantities so represented. The problems in dealing with integers (to select one example) on computers are of the following type: Is there enough storage to contain all the integers we need to deal with? Do we know a process that is certain to accomplish our goal on the integers stored in the computer? Have we removed the logical errors ("bugs") from the

---

Prof. Forsythe received his PhD at Brown University under W. Feller and J. D. Tamarkin. He was an instructor at Stanford, worked in meteorology with the Air Force and at UCLA, and worked in numerical analysis at Boeing Airplane Co., the Institute for Numerical Analysis, and at UCLA. He has been at Stanford since 1957 in mathematics and in computing science. He spent 1955–56 at the Courant Institute and 1966–67 at various computer centers in Europe, Asia, and Australia.

He is known for his extensive writing on random variables, meteorology, and computer science. In 1969 he received an MAA Lester Ford Award. His books are: *Dynamic Meteorology* (with J. Holmboe and W. Gustin, Wiley, 1945), *Bibliography of Russian Mathematics Books* (Chelsea, 1956), *Finite Difference Methods for Partial Differential Equations* (with W. R. Wasow, Wiley, 1960), and *Computer Solution of Linear Algebraic Systems* (with C. B. Moler, Prentice-Hall, 1967). *Editor.*

computer representation of this process? Is this the fastest possible process or, if not, does it operate quickly enough for us to get (and pay for) the answers we want?

The above problems are not trivial; there are surely pitfalls in dealing with them; and it is questionable whether math books suffice for their treatment. But they are not the subject of this paper. This paper is concerned with the simulated solution on a digital computer of the problems of algebra and analysis dealing with real and complex numbers. Such problems occur everywhere in applied science—for example, whenever it is required to solve a differential equation or a system of algebraic equations.

There are four properties of computers that are relevant to their use in the numerical solution of problems of algebra and analysis. These properties are causes of many pitfalls:

(i) Computers use not the real number system, but instead a simulation of it called a "floating-point number system." This introduces the problem of *round-off*.

(ii) The speed of computer processing permits the solution of very large problems. And frequently (but not always) large problems have answers that are much more *sensitive* to perturbations of the data than small problems are.

(iii) The speed of computer processing permits many more operations to be carried out for a reasonable price than were possible in the pre-computer era. As a result, the *instability* of many processes is conspicuously revealed.

(iv) Normally the intermediate results of a computer computation are hidden in the store of the machine, and never known to the programmer. Consequently the programmer must be able to detect errors in his process without seeing the warning signals of possible error that occur in desk computation, where all intermediate results are in front of the problem solver. Or, conversely, he must be able to prove that his process cannot fail in any way.

3. Floating-point number system. The badly named *real number system* is one of the triumphs of the human mind. It underlies the calculus and higher analysis to such a degree that we may forget how impossible it is to deal with real numbers in the real world of finite computers. But, however much the real number system simplifies analysis, practical computing must do without it.

Of all the possible ways of simulating real numbers on computers, one class is most widely used today—the *floating-point number system.* Here a number base β is selected, usually 2, 8, 10, or 16. A certain integer s is selected as the number of significant digits (to base β) in a computer number. An integer exponent e is associated with each nonzero computer number, and e must lie in a fixed range, say

$$m \leqq e \leqq M.$$

Finally, there is a sign $+$ or $-$ for each nonzero floating-point number.

Let $F=F(\beta, s, m, M)$ be the floating-point number system. Each nonzero

$x \in F$ has the base-β representation

$$x = \pm\ .d_1 d_2 \cdots d_s \cdot \beta^e,$$

where the integers $d_1, \cdots, d_s$ have the bounds

$$\begin{aligned} 1 &\leqq d_1 \leqq \beta - 1, \\ 0 &\leqq d_i \leqq \beta - 1 \qquad (i = 2, \cdots, s). \end{aligned}$$

Finally, the number 0 belongs to F, and is represented by

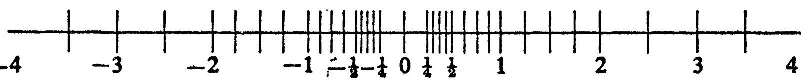
$$+.00 \cdots 0 \cdot \beta^m.$$

Actual computer number systems often differ in detail from the ideal one discussed here, but the differences are of only secondary relevance for the fundamental problems of round-off.

Typical floating-point systems in use correspond to the following values of the parameters:

$$\begin{aligned}
\beta &= \ \ 2,\ s = 48,\ m = -\ 975,\ M = 1071 \text{ (Control Data 6600)} \\
\beta &= \ \ 2,\ s = 27,\ m = -\ 128,\ M = \ \ 127 \text{ (IBM 7090)} \\
\beta &= 10,\ s = \ \ 8,\ m = -\ \ \ 50,\ M = \ \ \ \ 49 \text{ (IBM 650)} \\
\beta &= \ \ 8,\ s = 13,\ m = -\ \ \ 51,\ M = \ \ \ \ 77 \text{ (Burroughs 5500)} \\
\beta &= 16,\ s = \ \ 6,\ m = -\ \ \ 64,\ M = \ \ \ \ 63 \text{ (IBM System/360)} \\
\beta &= 16,\ s = 14,\ m = -\ \ \ 64,\ M = \ \ \ \ 63 \text{ (IBM System/360).}
\end{aligned}$$

Any one computer may be able to store numbers in more than one system. For example, the IBM System/360 uses the last two base-16 floating-point systems for scientific work, and also a certain base-10 system for accounting purposes.



F is not a continuum, nor even an infinite set. It has exactly $2(\beta-1)\beta^{s-1} \cdot (M-m+1)+1$ numbers in it. These are not equally spaced throughout their range, but only between successive powers of β and their negatives. The accompanying figure, reproduced from [3] by permission, shows the 33-point set F for the small illustrative system $\beta=2$, $s=3$, $m=-1$, $M=2$.

Because F is a finite set, there is no possibility of representing the continuum of real numbers in any detail. Indeed, real numbers in absolute value larger than the maximum member of F cannot be said to be represented at all. And, for many purposes, the same is true of nonzero real numbers smaller in magnitude than the smallest positive number in F. Moreover, each number in F has to represent a whole interval of real numbers. If x and y are two real numbers in

the range of F, they will usually be represented by the same number in F whenever $|x-y|/|x| \leqq \frac{1}{2}\beta^{-s}$; it is not important to be more precise here.

As a model of the real number system R, the set F has the arithmetic operations defined on it, as carried out by the digital computer. Suppose x and y are floating-point numbers. Then the true sum $x+y$ will frequently not be in F. (For example, in the 33-point system illustrated above let $x=5/4$ and $y=3/8$.) Thus the operation of addition, for example, must itself be simulated on the computer by an approximation called *floating-point addition*, whose result will be denoted by $\mathrm{fl}(x+y)$. Ideally, $\mathrm{fl}(x+y)$ should be that member of F which is closest to the true $x+y$ (and either one, in case of a tie). In most computers this ideal is almost, but not quite, achieved. Thus in our toy 33-point set F we would expect that $\mathrm{fl}(5/4+3/8)$ would be either $3/2$ or $7/4$. The difference between $\mathrm{fl}(x+y)$ and $x+y$ is called the *rounding error* in addition.

The reason that $5/4+3/8$ is not in the 33-point set F is related to the spacing of the members of F. On the other hand, a sum like $7/2+7/2$ is not in F because 7 is larger than the largest member of F. The attempt to form such a sum on most machines will cause a so-called *overflow signal*, and often the computation will be curtly terminated, for it is considered impossible to provide a useful approximation to numbers beyond the range of F.

While quite a number of the sums $x+y$ (for x, y in F) are themselves in F, it is quite rare for the true product $x\cdot y$ to belong to F, since it will always involve $2s$ or $2s-1$ significant digits. Thus the simulated multiplication operation, $\mathrm{fl}(x\cdot y)$, involves rounding even more often than floating addition. Moreover, overflow is much more probable in a product. Finally, the phenomenon of *underflow* occurs in floating-point multiplication, when two nonzero numbers x, y have a nonzero product that is smaller in magnitude than the smallest nonzero number in F. (Underflow is also possible, though unusual, in addition.)

The operations of floating-point addition and multiplication are commutative, but not associative, and the distributive law fails for them also. Since these algebraic laws are fundamental to mathematical analysis, working with floating-point operations is very difficult for mathematicians. One of the greatest mathematicians of the century, John von Neumann, was able to collaborate in some large analyses with floating-point arithmetic (see [**10**]), but they were extremely ponderous. Even his genius failed to discover a method of avoiding nonassociative analysis. Such a new method, called *inverse error analysis*, owes its origins to Cornelius Lanczos and Wallace Givens, and has been heavily exploited by J. H. Wilkinson. A detailed study of inverse error analysis is part of the subject of numerical analysis. We will mention it again in Section 5.

4. Two examples of round-off problems. One of the commonest functions in analysis is the exponential function e^x. Since it is so much used, it is essential to be able to have the value of e^x readily available in a computer program, for any (not too large or small) floating-point number x. There is nowhere near enough storage to file a table of all values of e^x, so one must instead have an

algorithm for recomputing e^x whenever it is needed. (By an algorithm we mean a discrete process that is completely defined and guaranteed to terminate.) There are, in fact, a great many different methods such an algorithm could use, and most scientific computing systems include such an algorithm. But let us assume such an algorithm did not exist on your computer, and ask how you would program it. This is a realistic model of the situation for a more obscure transcendental function of analysis.

Recall that, for any real (or complex) value of x, we can represent e^x by the sum of the universally convergent infinite series

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots.$$

Since you learned mathematics because it is useful, you might expect to use the series to compute e^x. Suppose—just for illustration—that your floating-point number system F is characterized by $\beta = 10$ and $s = 5$. Let us use the series for $x = -5.5$, as proposed by Stegun and Abramowitz [13]. Here are the numbers we get:

$$\begin{array}{rr} e^{-5.5} \approx & 1.0000 \\ & -\ 5.5000 \\ & +15.125 \\ & -27.730 \\ & +38.129 \\ & -41.942 \\ & +38.446 \\ & -30.208 \\ & +20.768 \\ & -12.692 \\ & +\ 6.9803 \\ & -\ 3.4902 \\ & +\ 1.5997 \\ & \vdots \\ \hline & +\ 0.0026363 \end{array}$$

(The symbol "$\approx$" means "equals approximately".) The sum is terminated when the addition of further terms stops changing it, and this turns out to be after 25 terms. Is this a satisfactory algorithm? It may seem so, but in fact $e^{-5.5} \approx 0.00408677$, so that the above series gets an answer correct to only about 36 percent! It is useless.

What is wrong? Observe that there has been a lot of cancellation in forming the sum of this alternating series. Indeed, the four leading (i.e., most significant)

digits of the eight terms that exceed 10 in modulus have all been lost. Professor D. H. Lehmer calls this phenomenon *catastrophic cancellation*, and it is fairly common in badly conceived computations. However, as Professor William Kahan has observed, this great cancellation is not the *cause* of the error in the answer—it merely *reveals* the error. The error had already been made in that the terms like 38.129, being limited to 5 decimal digits, can have only one digit that contributes to the precision of the final answer. It would be necessary for the term $(-5.5)^4/4!$ to be carried to 8 decimals (i.e., 9 leading digits) for it to include all 6 leading digits of the answer. Moreover, a tenth leading digit would be needed to make it likely that the fifth significant digit would be correct in the sum. The same is true of all terms over 10 in magnitude.

While it is usually possible to carry extra digits in a computation, it is always costly in time and space. For this particular problem there is a much better cure, namely, compute the sum for $x=5.5$ and then take the reciprocal of the answer:

$$\begin{aligned} e^{-5.5} &= 1/e^{5.5} \\ &= 1/(1 + 5.5 + 15.125 + \cdots) \\ &\approx 0.0040865, \text{ with our 5-decimal arithmetic.} \end{aligned}$$

With this computation, the error is reduced to 0.007 percent.

Note how much worse the problem would be if we wanted to compute e^x for $x=-100$.

Actual computer algorithms for calculating e^x usually use a rational function of x, for x on an interval like $0 \leqq x \leqq 1$. For x outside this interval, well-known properties of the exponential function are used to obtain the answer from the rational approximation to e^y, where $y = x - [x]$. The creation of such algorithms for special functions is a branch of numerical analysis in which the general mathematician can hardly be an expert. On the other hand, it is part of the author's contention that mathematics books ought to mention the fact that a Taylor's series is often a very poor way to compute a function.

I shall briefly state a second example. Recall from the calculus that

$$\text{(1)} \qquad \int_a^b \frac{dx}{x^p} = \frac{x^{1-p}}{1-p}\bigg]_a^b = \frac{1}{1-p}(b^{1-p} - a^{1-p}) \qquad (p \neq 1).$$

Now using a floating-point system with $\beta = 10$ and $s = 6$, let us evaluate the above formula for $a = 1$, $b = 2$, and $p = 1.0001$. We have

$$\text{(2)} \qquad I = \int_1^2 \frac{dx}{x^{1.0001}} = \frac{1 - 2^{-.0001}}{0.0001}.$$

If we use 6-place logarithms to evaluate $2^{-.0001}$, we have

$$\begin{aligned} \log_{10} 2 &\approx 0.301030, \\ \log_{10} 2^{-.0001} &\approx -0.0000301030 = -1 + 0.999970, \end{aligned}$$

whence, using our logarithm table again,

$$2^{-.0001} \approx 0.999930.$$

Thus, from (2), we get $I \approx 0.7$, an answer correct to only one digit.

The precise meaning of the restriction to $\beta = 10$, $s = 6$ is not so clear in the evaluation of $2^{-.0001}$ as it would have been in the previous example. However, the example does illustrate the fact that formula (1), which is precisely meaningful for real numbers as long as $p \neq 1$, is difficult to use with finite-precision arithmetic for p close to 1. Thus practical computation cannot admit the precise distinction between equality and inequality basic to pure mathematics. There are degrees of uncertainty caused by approximate equality.

5. Solving quadratic equations. The two examples of Section 4 were taken from the calculus. But we don't have to learn college mathematics to find algorithms. In ninth grade there is a famous algorithm for solving a quadratic equation, implicit in the following mathematical theorem:

THEOREM. *If a, b, c are real and $a \neq 0$, then the equation $ax^2 + bx + c = 0$ is satisfied by exactly two values of x, namely*

$$x_1 = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \tag{3}$$

and

$$x_2 = \frac{-b - \sqrt{b^2 - 4ac}}{2a}. \tag{4}$$

Let us see how these formulas work when used in a straightforward manner to induce an algorithm for computing x_1 and x_2. This time we shall use a floating-point system with $\beta = 10$, $s = 8$, $m = -50$, $M = 50$; this has more precision than many widely used computing systems.

CASE 1: $a = 1, b = -10^5, c = 1.$

The true roots of the corresponding quadratic equation, correctly rounded to 11 significant decimals, are:

$$x_1 \approx 99999.999990 \quad \text{(true)}$$

$$x_2 \approx 0.000010000000001 \text{ (true).}$$

If we use the expressions of the theorem, we compute

$$x_1 \approx 100000.00 \quad \text{(very good)}$$

$$x_2 \approx 0 \quad \text{(100 percent wrong).}$$

(The reader is advised to be sure he sees how x_2 becomes 0 in this floating-point computation.)

Once again, in computing x_2 we have been a victim of catastrophic cancellation, which, as before, merely reveals the error we made in having chosen this way of computing x_2. There are various alternate ways of computing the roots of a quadratic equation that do not force such cancellation. One of them follows from the easily proved formulas, true if $abc \neq 0$:

$$x_1 = \frac{2c}{-b - \sqrt{b^2 - 4ac}}, \tag{5}$$

$$x_2 = \frac{2c}{-b + \sqrt{b^2 - 4ac}}. \tag{6}$$

Now, if $b < 0$, there is cancellation in (4) and (5) but not in (3) and (6). And, if $b > 0$, there is cancellation in (3) and (6), but not in (4) and (5). Special attention must be paid to cases where b or c is 0.

At this point I should like to propose the following criterion of performance of a computer algorithm for solving a quadratic equation. This is stated rather loosely here, but a careful statement will be found in [2].

We define a complex number z to be *well within the range of F* if either $z = 0$ or

$$\beta^{m+2} \leqq |\operatorname{Re}(z)| \leqq \beta^{M-2} \quad \text{and}$$
$$\beta^{m+2} \leqq |\operatorname{Im}(z)| \leqq \beta^{M-2}.$$

This means that the real and imaginary parts of z are safely within the magnitudes of numbers that can be closely approximated by a member of F. The arbitrary factor β^2 is included as a margin of safety.

Suppose a, b, c are all numbers in F that are well within the range of F. Then they must be acceptable as input data to the quadratic equation algorithm. If $a = b = c = 0$, the algorithm should terminate with a message signifying that all complex numbers satisfy the equation $ax^2 + bx + c = 0$. If $a = b = 0$ and $c \neq 0$, then the algorithm should terminate with a message that no complex number satisfies the equation.

Otherwise, let z_1 and z_2 be the exact roots of the equation, so numbered that $|z_1| \leqq |z_2|$. (If $a = 0$, set $z_2 = \infty$.) Whenever z_1 is well within the range of F, the algorithm should determine a close approximation to z_1, in the sense of differing by not more than, say, $\beta + 1$ units in the least significant digit of the root.

The same should be done for z_2.

If either or both of the roots z_i are not well within the range of F, then an appropriate message should be given and the root (if any) that is well within the range of F should be determined to within a close approximation.

That concludes the loose specification of the desired performance of a quadratic equation solving algorithm. Let us return to a consideration of some typical equations, to see how the quadratic formulas work with them.

CASE 2: $a = 6$, $b = 5$, $c = -4$.

There is no difficulty in computing $x_1 \approx 0.50000000$ and $x_2 \approx -1.3333333$, or nearly these values, by whatever formula is used.

CASE 3: $a=6\cdot 10^{30}$, $b=5\cdot 10^{30}$, $c=-4\cdot 10^{30}$.

Since the coefficients in Case 3 are those of Case 2, all multiplied by 10^{30}, the roots are unchanged. However, application of any of the formulas (3)–(6) causes overflow to occur very soon, since $b^2>10^{50}$, out of the range of F. Probably this uniform large size of $|a|$, $|b|$, $|c|$ could be detected before entering the algorithm, and all three numbers could be divided through by the factor 10^{30} to reduce the problem to Case 2.

CASE 4: $a=10^{-30}$, $b=-10^{30}$, $c=10^{30}$.

Here z_1 is near 1, while z_2 is near 10^{60}. Thus our algorithm must determine z_1 very closely, even though z_2 is out of the range of F. Obviously any attempt to bring the coefficients to approximate equality in magnitude by simply dividing them all by the same number is doomed to failure, and might itself cause an overflow or underflow. This equation is, in fact, a severe test for a quadratic equation solver and even for the computing system in which the solver is run.

The reader may think that a quadratic equation with one root out of the range of F and one root within the range of F is a contrived example of no practical use. If so, he is mistaken. In many iterative algorithms which solve a quadratic equation as a subroutine, the quadratics have a singular behavior in which $a\to 0$ as convergence occurs. One such example is Muller's method **[9]** for finding zeros of general smooth functions of z.

CASE 5: $a=1.0000000$, $b=-4.0000000$, $c=3.9999999$.

Here the two roots are $z_1\approx 1.999683772$, $z_2\approx 2.000316228$. But applying the quadratic formulas (3), (4) gives

$$z_1 = z_2 = 2.0000000,$$

with only the first four digits correct. These roots fail badly to meet my criteria, but the difficulty here is different from that in the other examples. The equation corresponding to Case 5 is the first of our equations in which a small relative change in a coefficient a, b, c induces a much larger relative change in the roots z_1, z_2. This is a form of instability in the equation itself, and not in the method of solving it. To see how unstable the problem is, the reader should show that the computed roots 2.0000000 are the exact roots of the equation

$$0.999999992x^2 - 3.999999968x + 3.999999968 = 0,$$

in which the three coefficients differ, respectively, from the true a, b, c of Case 5 by less than one unit in the last significant digit. In this sense one can say that 2, 2 are pretty good roots for Case 5.

This last way of looking at rounding errors is called the *inverse error approach* and has been much exploited by J. H. Wilkinson. In general, it is characterized by asking how little a change in the data of a problem would be necessary to cause the computed answers to be the exact solution of the changed problem. The more intuitive way of looking at round off, the *direct error approach*, simply asks how wrong the answers are as solutions of the problems with its given data. While both methods are useful, the important feature of inverse error analysis is

that in many large matrix or polynomial problems, it can permit us easily to continue to use associative operations, and this is often very difficult with direct error analysis.

Despite the elementary character of the quadratic equation, it is probably still true that not more than five computer algorithms exist anywhere that meet the author's criteria for such an algorithm. Creating such an algorithm is not a very deep problem, but it does require attention to the goal and to the details of attaining the goal. It illustrates the sort of place that an undergraduate mathematics or computer science major can make a substantial contribution to computer libraries.

I wish to acknowledge that the present section owes a great deal to lectures by Professor William Kahan of the University of California, Berkeley, given at Stanford in the Spring of 1966.

6. Solving linear systems of equations. As the high school student moves from ninth grade on to tenth or eleventh, he will encounter the solution of systems of linear algebraic equations by Gauss' method of eliminating unknowns. With a little systematization, it becomes another algorithm for general use. I would like to examine it in the simple case of two equations in two unknowns, carried out on a computer with $\beta=10$, $s=3$.

Let the equation system be one treated by Forsythe and Moler [**3**]:

$$\begin{cases} 0.000100x + 1.00y = 1.00 \\ 1.00x \qquad\quad + 1.00y = 2.00. \end{cases} \tag{7}$$

The true solution, rounded correctly to the number of decimals shown, is

$$x \approx 1.00010, y \approx 0.99990 \quad \text{(truly rounded)}.$$

The Gauss elimination algorithm uses the first equation (if possible) to eliminate the first variable, x, from the second equation. Here this is done by multiplying the first equation by 10000 and then subtracting it from the second equation. When we work to three significant digits, the resulting system takes the form

$$\begin{cases} 0.000100x + 1.00y = 1.00 \text{ (the old first equation)} \\ \qquad\qquad - 10000y = -10000. \end{cases}$$

For just two equations, this completes the elimination of unknowns. Now commences the *back solution*. One solves the new second equation for y, finding that $y\approx 1.00$. This value is substituted into the first equation, which is then solved for x. One then finds $x\approx 0.00$. In summary, we have found

$$\begin{cases} y \approx 1.00 \\ x \approx 0.00. \end{cases}$$

Of course, this is awful! What went wrong? There was certainly no long accumulation of round-off errors, such as might be feared in a large problem. Nor is the

original problem unstable of itself, as it would be if the lines represented by the two equations (7) were nearly parallel.

There is one case in which it is impossible to eliminate x from the second equation—when the coefficient of x in the first equation is exactly 0. Were such an exact 0 to occur, the Gauss algorithm is preceded by interchanging the equations. Now, once again, if an *exact* zero makes a mathematical algorithm impossible, we should expect that a *near* zero will give a floating-point algorithm some kind of difficulty. That is a sort of philosophical principle behind what went wrong. And, in fact, the division by the nearly zero number 0.0001 introduced some numbers (10000) that simply swamped the much smaller, but essential, data of the second equation. That is what went wrong.

How could this be avoided? The answer is simple, in this case. If it is essential to interchange equations when a divisor is actually zero, one may suspect that it would be important, or at least safer, to interchange them when the coefficient of x in the first equation is much smaller in magnitude than the coefficient of x in the second equation. A careful round-off analysis given by J. H. Wilkinson [14] proves this to be the case, and good linear equation solvers will make the interchange whenever necessary to insure that the largest coefficient of x (in magnitude) is used as the divisor. Thus the elimination yields the system

$$\begin{cases} 1.00x + 1.00y = 2.00 \\ \qquad\quad\; 1.00y = 1.00. \end{cases}$$

After the back solution we find

$$\begin{cases} y \approx 1.00 \\ x \approx 1.00, \end{cases}$$

a very fine result.

This algorithm, with its interchanges, can be extended to n equations in n unknowns, and is a basic algorithm found in most computing centers.

The following example shows that there remains a bit more to the construction of a good linear equation solver. Consider the system

$$\begin{cases} 10.0x + 100000y = 100000 \\ 1.00x + 1.00y = 2.00. \end{cases} \tag{8}$$

If we follow the above elimination procedure, we see that interchanging the equations is not called for, since $10.0 > 1.00$. Thus one multiplies the first equation by 0.100 and subtracts it from the second. One finds afterwards, still working with $\beta = 10$, $s = 3$, that

$$\begin{cases} 10.0x + 100000y = 100000 \\ \qquad\quad - 10000y = -10000. \end{cases}$$

Back solving, one finds

$$\begin{cases} y \approx 1.00 \\ x \approx 0.00! \end{cases}$$

This is just as bad as before, for system (8) has the same solution as (7). Indeed, system (8) is easily seen to be identical with (7), except that the first equation has been multiplied through by 100000.

So the advice to divide by the largest element in the column of coefficients of x is not satisfactory for an arbitrary system of equations. What seems to be wrong with the system (8) is that the first equation has coefficients that are too large for the problem. Before entering the Gaussian elimination algorithm with interchanges, it is necessary to scale the equations so that the coefficients are roughly of the same size in all equations. This concept of scaling is not completely understood as yet, although in most practical problems we are able to do it well enough.

If you were faced with having to solve a nonsingular system of linear algebraic equations of order 26, for example, you might wonder how to proceed. Some mathematics books express the solution by Cramer's rule, in which each of the 26 components is the quotient of a different numerator determinant by a common denominator determinant. If you looked elsewhere, you might find that a determinant of order 26 is the sum of 26! terms, each of which is the product of 26 factors. If you decide to proceed in this manner, you are going to have to perform about $25 \cdot 26!$ multiplications, not to mention a similar number of additions. On a fast contemporary machine, because of the time required to do preparatory computations, you would hardly perform more than 100,000 multiplications per second. And so the multiplications alone would require about 10^{17} years, if all went well. The round-off error would usually be astronomical.

In fact, the solution can be found otherwise in about $(1/3) \cdot 26^2 \approx 5859$ multiplications and a like number of additions, and should be entirely finished in well under half a second, with very little round-off error. So it can pay to know how to solve a problem.

I wish to leave you with the feeling that there is more to solving linear equations than you may have thought.

7. When do we have a good solution? Another example of a linear algebraic system has been furnished by Moler [8]:

$$\begin{cases} 0.780x + 0.563y - 0.217 = 0 \\ 0.913x + 0.659y - 0.254 = 0. \end{cases} \tag{9}$$

Someone proposes two different approximate solutions to (9), namely

$$(x_1, y_1) = (0.999, -1.001)$$

and

$$(x_2, y_2) = (0.341, -0.087).$$

Which one is better? The usual check is to substitute them both into (9). We obtain

$$\begin{cases} 0.780x_1 + 0.563y_1 - 0.217 = -0.001243 \\ 0.913x_1 + 0.659y_1 - 0.254 = -0.001572 \end{cases}$$

and

$$\begin{cases} 0.780x_2 + 0.563y_2 - 0.217 = -0.000001 \\ 0.913x_2 + 0.659y_2 - 0.254 = 0. \end{cases}$$

It seems clear that (x_2, y_2) is a better solution than (x_1, y_1), since it makes the *residuals* far smaller.

However, in fact the true solution is $(1, -1)$, as the reader can verify easily. Hence (x_1, y_1) is far closer to the true solution than (x_2, y_2)!

A persistent person may ask again: which solution is *really* better? Clearly the answer must depend on one's criterion of goodness: a small residual, closeness to the true solution, or perhaps something else. Surely one will want different criteria for different problems. The pitfall to be avoided here is the belief that all such criteria are necessarily satisfied, if one of them is.

8. Sensitivity of certain problems. We now show that certain computational problems are surprisingly sensitive to changes in the data. This aspect of numerical analysis is independent of the floating-point number system.

We first consider the zeros of polynomials in their dependence on the coefficients. In Case 5 of Section 4 above, we noted that, while the polynomial x^2-4x+4 has the double zero 2, 2, the rounded roots of the polynomial equation

$$x^2 - 4x + 3.9999999 = 0 \tag{10}$$

are 1.999683772 and 2.000316228. Thus the change of just one coefficient from 4 to 3.9999999 causes both roots to move a distance of approximately .000316228. The displacement in the root is about 3162 times as great as the displacement in the coefficient.

The instability just described is a common one, and results from the fact that the square root of a small ϵ is far larger than ϵ. For the roots of (10) are the roots of

$$(x-2)^2 = \epsilon, \qquad \epsilon = .0000001,$$

and these are clearly $2 \pm \sqrt{\epsilon}$. For equations of higher degree, a still more startling instability would have been possible.

However, it is not only for polynomials with nearly multiple zeros that instability can be observed. The following example is due to Wilkinson [**14**]. Let

$$\begin{aligned} p(x) &= (x-1)(x-2)\cdots(x-19)(x-20) \\ &= x^{20} - 210x^{19} + \cdots. \end{aligned}$$

The zeros of $p(x)$ are $1, 2, \cdots, 19, 20$, and are well separated. This example

evolved at a place where the floating-point number system had $\beta=2$, $s=30$. To enter a typical coefficient into the computer, it was necessary to round it to 30 significant base-2 digits. Suppose that a change in the 30-th most significant base-2 digit is made in *only one* of the twenty coefficients. In fact, suppose that the coefficient of x^{19} is changed from -210 to $-210-2^{-23}$. How much effect does this small change have on the zeros of the polynomial?

To answer this, Wilkinson carefully computed (using $\beta=2$, $s=90$) the roots of the equation $p(x)-2^{-23}x^{19}=0$. These are now listed, correctly rounded to the number of digits shown:

1.00000 0000	10.09526 6145 $\pm$ 0.64350 0904i
2.00000 0000	11.79363 3881 $\pm$ 1.65232 9728i
3.00000 0000	13.99235 8137 $\pm$ 2.51883 0070i
4.00000 0000	16.73073 7466 $\pm$ 2.81262 4894i
4.99999 9928	19.50243 9400 $\pm$ 1.94033 0347i
6.00000 6944	
6.99969 7234	
8.00726 7603	
8.91725 0249	
20.84690 8101	

Note that the small change in the coefficient -210 has caused ten of the zeros to become complex, and that two have moved more than 2.81 units off the real axis! Of course, to enter $p(x)$ completely into the computer would require many more roundings, and actually computing the zeros could not fail to cause still more errors. The above table of zeros was produced by a very accurate computation, and does not suffer appreciably from round-off errors. The reason these zeros moved so far is not a round-off problem—it is a matter of sensitivity. Clearly zeros of polynomials of degree 20 with well-separated zeros can be much more sensitive to changes in the coefficients than you might have thought.

To motivate a second example, let me quote a standard theorem of algebra: *In the ring of square matrices of fixed order n, if $AX=I$, where I is the identity matrix of order n, then $XA=I$.*

It follows from this theorem and continuity considerations that, if A is a fixed matrix and X a variable one, and if $AX-I\to\theta$, the zero matrix, then also $XA-I\to\theta$. Hence, if $AX-I$ is small in some sense, then $XA-I$ is also small. However, as with polynomials, one's intuition may not be very good at guessing how small these smallnesses are. Here is an example: Fix

$$A=\begin{bmatrix} 9999 & 9998 \\ 10000 & 9999 \end{bmatrix}.$$

Let

$$X = \begin{bmatrix} 9999.9999 & -9997.0001 \\ -10001 & 9998 \end{bmatrix}.$$

Then a computation without round-off shows that

$$AX - I = \begin{bmatrix} .001 & .0001 \\ 0 & 0 \end{bmatrix}.$$

From the last equality the reader may conclude that X is close, though not equal, to the unique inverse A^{-1}. However, another calculation without round-off shows that

$$XA - I = \begin{bmatrix} 19997.0001 & 19995.0003 \\ -19999 & -19995 \end{bmatrix}.$$

Thus the quantities $AX-I$ and $XA-I$, which must vanish together, can be of enormously differing magnitudes in a sensitive situation, even for matrices of order 2.

The true inverse matrix is given by

$$A^{-1} = \begin{bmatrix} 9999 & -9998 \\ -10000 & 9999 \end{bmatrix},$$

and this is hardly close to X.

9. A least-squares problem of Hilbert. The following least-squares problem was discussed by the great mathematician David Hilbert [6], and leads to some interesting matrices. Fix $n \geqq 1$. Let $f(t)$ be given and continuous for $0 \leqq t \leqq 1$. We wish to approximate $f(t)$ as well as we can by a polynomial $x_1 + x_2 t + x_3 t^2 + \cdots + x_n t^{n-1}$ of degree $n-1$. To be more precise, we wish to determine $x_1, x_2, \cdots, x_n$ so that

$$\Phi(x) = \int_0^1 [f(t) - x_1 - x_2 t - \cdots - x_n t^{n-1}]^2 dt$$

is as small as possible. It is not difficult to show that the minimizing vector of coefficients x exists, is unique, and can be determined by solving the system of n simultaneous equations

$$\frac{\partial \Phi}{\partial x_i} = 0 \qquad (i = 1, 2, \cdots, n). \tag{11}$$

If you carry out the algebra, you find that (11) is equivalent to the system of n linear algebraic equations

$$Ax = b, \tag{12}$$

where

$$a_{i,j} = \int_0^1 t^{i-1} t^{j-1} dt = \frac{1}{i + j - 1} \qquad (i, j = 1, 2, \cdots, n) \tag{13}$$

and

$$b_i = \int_0^1 t^{i-1} f(t) dt \qquad (i = 1, 2, \cdots, n). \tag{14}$$

The matrix A of coefficients in (12) is now called the *Hilbert matrix* (*of order* n), and is denoted by H_n:

$$H_n = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \cdots & \frac{1}{n} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \cdots & \frac{1}{n+1} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \frac{1}{n} & \frac{1}{n+1} & \frac{1}{n+2} & \cdots & \frac{1}{2n-1} \end{bmatrix}$$

The equations (12) with matrix $A = H_n$ are called the *normal equations* for this least-squares problem. It appears that all one has to do is to find and use a quadrature rule for approximating the b_i in (14), and then solve the system (12). This is certainly the standard advice in books on practical statistics.

However, what is observed is that for n bigger than 8 or 9 (the threshold depends on the system used), programs for solving linear equations in ordinary floating-point precision are simply unable to solve (12). Moreover, for problems that can be solved (say $n=6$), there are enormous differences in the solution vectors x for apparently identical problems on slightly different machines. Why all this trouble?

Let me try to explain the sensitivity of the problem first. Let $T_n = H_n^{-1}$. Then it can be proved that

$$T_6 = \begin{bmatrix} 36 & -630 & 3360 & -7560 & 7560 & -2772 \\ -630 & 14700 & -88200 & 211680 & -220500 & 83160 \\ 3360 & -88200 & 564480 & -1411200 & 1512000 & -582120 \\ -7560 & 211680 & -1411200 & 3628800 & -3969000 & 1552320 \\ 7560 & -220500 & 1512000 & -3969000 & 4410000 & -1746360 \\ -2772 & 83160 & -582120 & 1552320 & -1746360 & 698544 \end{bmatrix}.$$

This means that a change of 10^{-6} in just the one element b_5 will produce changes in the solution vector x of

$$(.00756,\ -\ .2205,\ 1.512,\ -\ 3.969,\ 4.41,\ -\ 1.74636)^T.$$

Such changes are unavoidable in a system with $\beta=10$ and $s=7$. This means that some of the coefficients of the best fitting polynomial of degree 5 will have unavoidable uncertainties of the order of 4 units. This may give some explanation of the instability in the answers. More details are in Section 19 of [3].

Here are approximate values of t_n, the maximum elements in T_n, for $n \leqq 10$;

n	t_n	n	t_n
2	$1.20 \cdot 10^1$	6	$4.41 \cdot 10^6$
3	$1.92 \cdot 10^2$	7	$1.33 \cdot 10^8$
4	$6.48 \cdot 10^3$	8	$4.25 \cdot 10^9$
5	$1.79 \cdot 10^5$	9	$1.22 \cdot 10^{11}$
		10	$3.48 \cdot 10^{12}$

It cannot be demonstrated here, but if $t_n \gg \beta^s$, you just cannot solve the system $H_n x = b$ with s-digit arithmetic in base β.

The conclusion of this example is that one should not follow a statistics book blindly here. It is much better to arrange things so that matrices of Hilbert type do not arise, even approximately. And when they do, one must be sure to use enough precision that $t_n \ll \beta^s$. There are other ways of attacking least-squares problems which are less sensitive to the data.

10. Instability in solving ordinary differential equations. The standard initial-value problem for a single ordinary differential equation $dy/dx = f(x, y)$ is to determine $y(x)$ as accurately as possible for $x > 0$, given $y(0)$. In one very common class of methods (the multistep methods) of solving this problem approximately, one picks a fixed interval $h > 0$, and determines y_n to approximate $y(nh)$ for $n = 1, 2, \cdots$. One highly recommended multistep method in desk-computing days was the *Milne-Simpson method.* Here one let $y_0 = y(0)$, the given initial value, and determined y_1 by some method not mentioned here. Let $y_n' = f(nh, y_n)$. The basic idea was to determine y_{n+1} from y_{n-1} and y_n $(n = 1, 2, \cdots)$ by the integral

$$y_{n+1} = y_{n-1} + \int_{(n-1)h}^{(n+1)h} f(x, y(x))dx. \tag{15}$$

Since the integral in (15) cannot usually be evaluated exactly, Milne's actual idea was to approximate it by Simpson's formula, and so let

$$y_{n+1} = y_{n-1} + \frac{h}{3}\,(y_{n-1}' + 4y_n' + y_{n+1}'). \tag{16}$$

At the time we seek to find y_{n+1} from (16) we know y_{n-1} and y_n, and hence y_{n-1}' and y_n'; but y_{n+1}' is not known. For general f, Milne [7] determined the solution of (16) by an iterative process that is irrelevant to the present discussion. Let

us merely assume that y_{n+1} has been found so that (16) holds, where $y'_{n+1} = f((n+1)h, y_{n+1})$, and that this has been done for $n=1, 2, \cdots$, as far as we wish to go. This method was highly recommended by Milne for solution of ordinary differential equations at a desk calculator, and it seemed to work very well indeed. Most problems were probably solved within 30 steps or less.

As soon as automatic digital computers arrived on the scene, users of the Milne-Simpson method started to find extraordinary behavior in certain problems. To illustrate what happened, let us take the very simple test problem

$$dy/dx = f(x, y) = -y, \quad \text{with } y(0) = 1.$$

The true solution, of course, is $y = e^{-x}$.

Take $h=0.1$, and carry out the Milne-Simpson process with $y_0 = 1$ and $y_1 = 0.90483742$, an 8-decimal correctly rounded value of $e^{-0.1}$. This is not something you can do in your head, and so I will give you the results, as computed on a system with $\beta = 10$, $s = 8$:

x	y_{computed}	e^{-x}
.2	.81873069	.81873075
.3	.74081817	.74081822
. . .	. . .	. . .
8.0	.00033519912	.00033546263
8.1	.00030380960	.00030353914
. . .	. . .	. . .
13.2	.00000036689301	.0000018506012
13.3	.0000032084360	.0000016744932
13.4	−.000000070769248	.0000015151441
. . .	. . .	. . .

We see that by $x=8.0$ a noticeable oscillation has set in, whereby successive values of y_n alternate in being too low and too high. By $x=13.4$ this oscillation has grown so violent that it has (for the first time) actually thrown the sign of y_n negative, which is unforgivable in anything simulating a real exponential function!

The Milne-Simpson method is very accurate, in that the Simpson formula is an accurate approximation to the above integral. What can be the matter?

Since $f(x, y) = -y$, we can explicitly write down the formula (16) in the form

$$y_{n+1} = y_{n-1} - \frac{h}{3}(y_{n-1} + 4y_n + y_{n+1}).$$

Thus the computed $\{y_i\}$ satisfy the 3-term recurrence relation

$$\left(1 + \frac{h}{3}\right) y_{n+1} + \frac{4h}{3} y_n - \left(1 - \frac{h}{3}\right) y_{n-1} = 0. \tag{17}$$

We know that the general solution of (17) takes the form

$$y_n = A_1 \lambda_1^n + A_2 \lambda_2^n, \tag{18}$$

where λ_1, λ_2 are the roots of

$$\left(1+\frac{h}{3}\right)\lambda^2+\frac{4h}{3}\lambda-\left(1-\frac{h}{3}\right)=0. \tag{19}$$

Some algebra and elementary analysis show that

$$\lambda_1 = 1 - h + O(h^2), \qquad \text{as } h \to 0,$$

$$\lambda_2 = -\left(1+\frac{h}{3}\right) + O(h^2), \qquad \text{as } h \to 0.$$

Putting these values of λ_1, λ_2 into (18), and using the relation $nh=x$, we find that, for small h,

$$\begin{aligned} y_n &\approx A_1(1-h)^n + (-1)^n A_2\left(1+\frac{h}{3}\right)^n \\ &= A_1(1-h)^{x/h} + (-1)^n A_2\left(1+\frac{h}{3}\right)^{(3/h)\cdot(x/3)} \\ &\approx A_1 e^{-x} + (-1)^n A_2 e^{x/3}. \end{aligned}$$

The first term is the desired solution if $A_1=1$, and the second is an unwelcome extra solution of the difference equation (17) of the Milne-Simpson method. Now the initial conditions might have been chosen exactly so that $A_1=1$ and $A_2=0$. (They were roughly of this nature.) Had they been so chosen, and if the solution could have proceeded without round-off error, the unwanted term in A_2 would never have appeared. But, in fact, a small amount of this solution was admitted by the initial condition, and some more of it crept in as the result of round-off. Then, after enough steps, the size of $e^{x/3}$ caused the unwanted term to dominate the solution, with its oscillating sign.

This disaster never occurred in desk computation, so far as we know, because at a desk one just doesn't carry out enough steps. Professor Milne has just told me, however, that he did occasionally observe harmless oscillations in the low-order digits.

The moral of this example is that not only are math books not enough, but even old numerical analysis books are not enough to keep you out of some pitfalls!

11. Instability in solving a partial differential equation. The following is a simple problem for the heat equation. Suppose a homogeneous insulated rod of length 1 is kept at temperature 0 at one end, and at temperature 1 at the other end. If the entire rod is initially at temperature 0, how does it warm up?

Let $u=u(x, t)$ denote the temperature at time t at that part of the rod that is x units from the cold end. Then, if the units are chosen to make the conductivity 1, the temperature u satisfies the differential equation

$$\frac{\partial^2 u}{\partial x^2} = \frac{\partial u}{\partial t} \quad (0 < x < 1; t > 0), \tag{20}$$

with end and initial conditions

$$\begin{cases} u(0, t) = 0 & (t > 0), \\ u(1, t) = 1 & (t > 0), \\ u(x, 0) = 0 & (0 < x < 1). \end{cases} \tag{21}$$

This problem can perhaps best be solved by separation of variables and Fourier series. But let us apply the method of finite differences, which might in any case be needed for a more difficult problem. To do this, we divide the length of the rod into equal intervals, each of length h. And we divide the time interval $[0, \infty)$ into equal intervals of length k. Instead of trying to determine $u(x, t)$ for all x and t, we shall limit ourselves to computing $u(x, t)$ on the discrete net of points of type (mh, nk), for integers m, n. The heat equation (20) can then be simulated by a number of finite-difference equations, of which we pick one:

$$\frac{u(x - h, t) - 2u(x, t) + u(x + h, t)}{h^2} = \frac{u(x, t + k) - u(x, t)}{k} \cdot \tag{22}$$

Equation (22) can be used to determine $u(x, t)$ for all net points in the infinite strip of the problem, as follows: Solve (22) for $u(x, t+k)$ in terms of $u(x-h, t)$, $u(x, t)$, $u(x+h, t)$. Thus compute $u(x, k)$ for $x = h, 2h, \cdots, (n-1)h$ in terms of the given initial conditions on the line $t = 0$. The given end conditions give $u(0, k)$ and $u(1, k)$. With this set of values of u at all points of the net with $t = k$, we can continue and compute all values on the net for $t = 2k$, etc. The computation is very attractive, because each new value of $u(x, t+k)$ is determined explicitly from (22)—there is no need to solve a system of simultaneous equations.

How does the solution behave? To try a case, we pick $h = 0.1$ and $k = 0.01$. Thus the rod is represented by 9 interior points and two endpoints, and we get a solution at time steps 0.01 apart. Just to show the behavior of the solution of (22), we give the value of the temperature $u(0.5, t)$ at the midpoint of the rod, computed with $\beta = 10$, $s = 8$, for selected times:

t	$u(0.5, t)$ computed from $k = 0.01$
0	0
$\cdots$	$\cdots$
0.05	1
0.06	-4
0.07	16
$\cdots$	$\cdots$
0.15	132276
$\cdots$	$\cdots$
0.20	-28157050
$\cdots$	$\cdots$
0.99	$+1.0196022 \cdot 10^{44}$
1.00	$-2.9590007 \cdot 10^{44}$.

The values in the table are ridiculous, of course. It is a classical example of instability. Common sense and mathematics both tell us that the real temperature can never get outside the range of $0 \leqq u(x, t) \leqq 1$. Our difference-equation problem is a disastrous model of the continuous problem, even though both difference expressions in (22) are reasonable models of the derivatives in (20).

This terrible pitfall has been known for at least 20 years, and yet new problem solvers keep on rediscovering it.

It is interesting to note that if one selects a time step only half as long, the computation proceeds very nicely. Here is the corresponding table of values of $u(0.5, t)$ for a computation ($\beta = 10$, $s = 8$) with $h = 0.1$, $k = 0.005$:

t	$u(0.5, t)$ computed for $k = 0.005$
0	0
. . .	. . .
0.05	.10937500
0.06	.14599609
0.07	.17956543
. . .	. . .
0.15	.35637261
. . .	. . .
0.20	.41304382
. . .	. . .
1.00	.49997173.

The values of the midpoint temperature appear to be converging to 0.5, as they obviously should in the physical problem.

What is the reason for the great difference in behavior between $k = 0.005$ and $k = 0.01$? The matter can be analyzed in many ways, and here is one simple approach. Let $\lambda = k/h^2$. Then, from (22),

$$u(x, t + k) = \lambda u(x - h, t) + (1 - 2\lambda) u(x, t) + \lambda u(x + h, t). \tag{23}$$

Hence, if $0 < \lambda \leqq 1/2$, the formula (23) represents $u(x, t+k)$ as a weighted average with nonnegative weights of $u(x-h, t)$, $u(x, t)$, and $u(x+h, t)$. Hence $u(x, t+k)$ will always be between the maximum and minimum values of $u(x, t)$. But if $\lambda > 1/2$, the weights alternate in sign and thus permit a solution in which

$$| u(x, t + k) | = \lambda | u(x - h, t) | + (2\lambda - 1) | u(x, t) | + \lambda | u(x + h, t) |.$$

Here the sum of the weights is $4\lambda - 1 > 1$. This permits an exponential growth of a solution with an alternating sign pattern.

Thus the condition $0 < \lambda = k/h^2 \leqq 1/2$ is essential to keep the solution bounded. A deeper discussion found, for example, in Forsythe and Wasow [4] proves that the solution of (22) converges to the solution of (20) uniformly for all (x, t) with $0 \leqq x \leqq 1$, $0 < t \leqq T < \infty$, as $h \to 0$, $k \to 0$ in such a way that $k/h^2 \leqq 1/2$.

The proof of convergence and an analysis of the stability of (22) can be carried out by means of Fourier analysis. The stability can be examined in more detail by studying the eigenvalues and eigenvectors of the linear transformation (23) that maps each line of solutions onto the next line.

Note that in our two tables we had $\lambda=1$ and $\lambda=1/2$, respectively.

12. Round-off errors in polynomial deflation. Our final example, due to Wilkinson [**14**], shows a more subtle effect of round-off error that arises in the course of finding polynomial zeros. The quartic polynomial

$$P_4(x) = x^4 - 6.7980x^3 + 2.9948x^2 - 0.043686x + 0.000089248$$

has zeros that, correctly rounded, are as follows:

$$0.0024532,\ 0.012576,\ 0.45732,\ 6.32565.$$

I. Suppose first that we somehow compute the zero 0.0024532, and then *deflate* P_4 to a cubic by dividing $P_4(x)$ by $x-0.0024532$, using $\beta=10$, $s=5$. If we do, the resulting cubic has zeros

$$0.012576,\ 0.457315,\ 6.32561,$$

so that the main error introduced by this deflation is a change of the largest zero by 4 units in its last place.

II. Suppose, on the other hand, that we first compute the zero 6.3256, and then deflate P_4 to a cubic by dividing $P_4(x)$ by $x-6.3256$, again using 5-place decimal arithmetic. If so, the resulting cubic has the zeros

$$0.0026261 \pm 0.064339i,\ 0.467148.$$

We have perturbed two of the remaining zeros beyond recognition, and have changed the second significant digit of the third!

Thus it appears to matter a great deal which zero of P_4 we locate first. For the present case we can get a feeling for what is happening by examining the process of division of $P_4(x)$ by the linear factors. We use detached coefficients:

First, the division by $x-0.0024532$:

$$\begin{array}{lllll}
1 & -\,6.7980 & +\,2.9948 & -\,0.043686 & +\,0.000089248 \\
 & -\,0.0024532 & +\,0.166707206 & -\,0.00730587492 & +\,0.000089247416 \\
\hline
1 & -\,6.7955 & +\,2.9781 & -\,0.036380. &
\end{array}$$

Thus the cubic that results from the first deflation is

$$\tilde{P}_3(x) = x^3 - 6.7955x^2 + 2.9781x - 0.036380.$$

Moreover, a careful examination of the division shows that $\tilde{P}_3(x)$ is *exactly* (i.e., without round-off) equal to the quotient of

$$\tilde{P}_4(x) = x^4 - 6.7979532x^3 + 2.9947707206x^2 - 0.04368587492x + 0.000089247416$$

by $x-0.0024532$. Hence the zeros of $\tilde{P}_3$ are exactly the zeros of $\tilde{P}_4$ except for 0.0024532. Note that all the coefficients of $\tilde{P}_4$ and P_4 are quite close, so it is reasonable to expect that the zeros of P_4 and $\tilde{P}_4$ should be close (as they are).

Now we show the deflation by $x-6.3256$:

$$\begin{array}{lllll} 1 - 6.7980 & + 2.9948 & - 0.043686 & + 0.000089248 \\ \quad - 6.3256 & + 2.98821344 & - 0.04174896 & + 0.0122526872 \\ \hline 1 - 0.4724 & + 0.0066 & - 0.001397. & \end{array}$$

Thus the result of this deflation is a cubic $\hat{P}_3(x) = x_3 - 0.4724x^2 + 0.0066x - 0.001397$. Again, $\hat{P}_3(x)$ is exactly the quotient of

$$\hat{P}_4(x) = x^4 - 6.7980x^3 + 2.99481344x^2 - 0.04368596x + 0.0122526872$$

by $x-6.3256$. Note that P_4 and $\hat{P}_4$ differ very much in their constant terms. Hence the product of the roots of $\hat{P}_4$ must be very different from that for P_4. This is an explanation for the great shift of the zeros of $\hat{P}_3$.

Further analysis shows that the shift in zeros during this kind of deflation is generally small when deflation is made with zeros of small modulus, and is generally large when deflation is based on zeros of large modulus. Thus it is better to get zeros of small modulus first in using a polynomial solver with deflation in the above manner.

Of course, any zero of a deflated polynomial can be refined by use of the original polynomial, and that is normally done. But, zeros that change as much as those above are difficult to refine, since the refinement process may converge to the wrong zero.

13. Conclusions. Around ten years ago, when I last read a number of them, most mathematics books that dealt with numerical methods at all were from ten to fifty years out of date. In the past ten years, many excellent new methods have been devised for most of the elementary problems—methods that are well adapted to automatic computers, and work well. Let me cite a few examples of important algorithms hardly known ten years ago:

1. For getting eigenvalues of stored square matrices, there is an excellent method that starts with the transformation of Householder (1958), and follows it with the QR-algorithm of Francis (1961–62) and Kublanovskaja (1961). It is the method of choice for most problems. For references, see Wilkinson [**15**].

2. For solving ordinary differential equations, special methods have been developed by Gear [**5**], Osborne [**11**], and many others which can deal with so-called *stiff* equations. (Roughly speaking, a stiff equation is one whose solutions contain very rapidly decaying transients which contribute nothing to the long-term solution, but which interfere drastically with most numerical methods of solving the equation.)

3. For evaluating the definite integral of a smooth function of one real variable, the method of Romberg (see vol. 2 of Ralston and Wilf [**12**]) has proved to be very useful.

4. For minimizing a smooth real-valued function of n real variables, a vari-

ant by Fletcher and Powell [1] of a method of Davidon is far superior to anything used in the 1950's. And there are still more recent methods.

Many other examples could be given. Indeed, the 1960's have proved almost explosive in the number of newly invented algorithms that have supplanted those known earlier. Of the methods known years ago for common numerical problems, only Gauss' systematic elimination method for solving linear algebraic equation systems with dense, stored matrices remains supreme today, and even it must be augmented with scaling and pivoting decisions, as we noted in Section 6 above. Newton's method for solving a nonlinear system of equations is still much used today, though it has strong competition from newer methods.

Because of my knowledge of mathematics texts ten years ago, and my knowledge of the explosive increase in numerical methods in the 1960's, I am confident that today's mathematics books cannot be trusted to include important knowledge about computer methods. As we noted in Section 10 above, you can't trust early numerical analysis textbooks either.

On the other hand, there are experts in numerical analysis. They have societies in which methods are presented and discussed. The Society for Industrial and Applied Mathematics (SIAM) and the Special Interest Group on Numerical Mathematics (SIGNUM) of the Association for Computing Machinery (ACM) are the most active in this country. There are a number of journals with important information. For a start, you might consult the keyword-in-context index of *Computing Reviews*, the review journal published by ACM, as well as the algorithms in the *Communications of the ACM* and in *Numerische Mathematik*. Modern monographs and textbooks in numerical analysis are slowly appearing, and the beginner might profitably consult Ralston and Wilf [12], especially volume 2.

It might be noted as a digression that, just as mathematics departments mainly ignore modern numerical analysis, so also the newly created computer science departments often give the subject little attention, since they are so busy with a variety of important nonnumerical fields. Thus numerical analysts remain a small corps of specialists whose greatest appreciation probably comes from the users of mathematical programs.

Students of mathematics are well equipped to read about numerical methods. Why should they repeat the classical blunders of generations past? Why aren't they informed of the existence of good numerical methods, and roughly where to find them?

Remembering that most students take mathematics in order to apply it on computers, I ask why mathematics courses shouldn't reflect a true awareness of how computing is done? Why shouldn't students demand in their mathematics courses a greater awareness of the points of contact of pure mathematics and its practice on a computer?

Of course, a mathematics instructor can shrug his shoulders and say that actual computing problems don't interest him, and suggest that his students contact a numerical analyst sometimes. If the instructor actually says this out

loud, it at least has the virtue that the students may realize immediately that the mathematics is not applicable directly instead of having to discover it for themselves. It still sounds irresponsible to me. After all, society has been supporting mathematicians pretty well for the past 25 years—not because mathematics is a beautiful art form, which it is—but because mathematics is useful, which it also is. But this would seem to imply that a mathematician should convey some awareness of the main ways in which his subject is used.

On the other hand, a mathematics course cannot really include very much numerical analysis. Wilkinson's treatise [15] on computing eigenvalues is 700 pages long, and can hardly be summarized in every course on linear algebra! As a practical matter, then, the mathematics instructor's main responsibility is to be aware of the main features of practical computing in the areas of his mathematics courses, and mention occasional points of contact, while giving his students pertinent references to important algorithmic materials in other books.

If one just ignores the relations between mathematics and its important applications, I fear that an instructor is running the risk of being exposed by some technological chapter of the Students for Democratic Society for not being relevant, and that is a very nasty accusation nowadays. Why risk it?

The preparation of this manuscript was supported in part by the Office of Naval Research (NR 044 211), the National Science Foundation (GJ 798), and the Atomic Energy Commission (Stanford PA #18). This material was presented by invitation to the Mathematical Association of America in Eugene, Oregon, 25 August 1969. Reproduction in whole or in part is permitted for any purpose of the United States Government.

References

1. R. Fletcher and M. J. D. Powell, A rapidly convergent descent method for minimization, Computer J., 6 (1963) 163–168.

2. George E. Forsythe, What is a satisfactory quadratic equation solver, pp. 53–71 of B. Dejon and P. Henrici (editors), Constructive Aspects of the Fundamental Theorem of Algebra, Wiley-Interscience, New York, 1969.

3. George E. Forsythe and Cleve B. Moler, Computer Solution of Linear Algebraic Systems, Prentice-Hall, Englewood Cliffs, N. J., 1967.

4. George E. Forsythe and Wolfgang R. Wasow, Finite-Difference Methods for Partial Differential Equations, Wiley, New York, 1960.

5. C. W. Gear, The automatic integration of stiff ordinary differential equations, pp. A81–A86 of Anonymous, Proceedings IFIP Congress 68, North Holland Publishing Co, Amsterdam, 1968.

6. D. Hilbert, Ein Betrag zur Theorie des Legendre'schen Polynoms, Acta Math., 18 (1894) 155–160.

7. William Edmund Milne, Numerical Solution of Differential Equations, Wiley, New York, 1953.

8. Cleve B. Moler, Numerical solution of matrix problems, pp. 15–26 of Anonymous, The Digest Record of the 1969 Joint Conference on Mathematical and Computer Aids to Design, I.E.E.E. Catalogue No. 69 C 63-C, 1969.

9. David E. Muller, A method for solving algebraic equations using an automatic computer, Math. Tables and Other Aids to Computation, 10 (1956) 208–215.

10. John von Neumann and H. H. Goldstine, Numerical inverting of matrices of high order, Bull. Amer. Math. Soc., 53 (1947) 1021–1099, and Proc. Amer. Math. Soc., 2 (1951) 188–202.

11. Michael R. Osborne, A new method for the integration of stiff systems of ordinary differential equations, pp. A86–A90 of Anonymous, Proceedings IFIP Congress 68, North Holland Publishing Co., Amsterdam, 1968.

12. Anthony Ralston and Herbert S. Wilf, Mathematical Methods for Digital Computers, Wiley, New York, Vol. 1, 1960, and Vol. 2, 1967.

13. Irene A. Stegun and Milton Abramowitz, Pitfalls in computation, J. Soc. Indust. Appl. Math., 4 (1956) 207–219.

14. J. H. Wilkinson, Rounding Errors in Algebraic Processes, Prentice-Hall, Englewood Cliffs, N. J., 1963.

15. ———, The Algebraic Eigenvalue Problem, Clarendon Press, Oxford, 1965.

ON LINEAR INEQUALITIES IN COMPLEX SPACE

R. N. KAUL, University of Delhi

1. Introduction. In [1] Eisenberg has given characterization of the set of all supports of a convex function. This extends the famous Farkas theorem [2, 3] and is useful in mathematical programming. The generalization of Farkas' theorem for a complex space has been obtained by Levinson [4]. Its extension, which is the object of the present note, is contained in the theorem that follows.

Let A be a complex $m\times n$ matrix. The conjugate transpose of A will be denoted by A^*. If u and v are two complex vectors with the same number of components, then

$$u\cdot v = v^* u = \sum_i u_i\bar{v}_i.$$

Let $a^{(j)}$ be the columns of the matrix A and ξ an $n\times 1$ complex vector. Then we write

$$A\xi = \sum_{j=1}^{n} a^{(j)}\xi_j = t.$$

If α is a real vector such that $0\leqq\alpha_j\leqq\pi/2$ for $j=1, 2, \cdots, n$, then by

$$|\arg t| \leqq \alpha, \qquad \text{resp.,}\ |\arg t| \leqq (\pi/2) - \alpha,$$

we mean that

$$|\arg t_j| \leqq \alpha_j, \qquad \text{resp.,}\ |\arg t_j| \leqq (\pi/2) - \alpha_j.$$

We now state the theorem:

THEOREM. *Let A be an $m\times n$ matrix with complex entries, C an $n\times n$ hermitian positive semi-definite matrix, and b a complex $n\times 1$ vector. Then*

$$(z^* C z)^{1/2} + \mathrm{Re}(b^* z) \geqq 0$$

for all $n \times 1$ *complex vectors* z *satisfying*

$$|\arg Az| \leqq \alpha \tag{1}$$

if and only if there exists an $m \times 1$ *vector* ξ *and an* $n \times 1$ *vector* $\mathfrak{n}$ *such that*

$$A^*\xi = b + C\mathfrak{n} \tag{2}$$

$$|\arg \xi| \leqq (\pi/2) - \alpha, \tag{3}$$

and $\mathfrak{n}$ *satisfies* (1) *and*

$$\mathfrak{n}^* C\mathfrak{n} \leqq 1. \tag{4}$$

The above theorem is used as the principal tool for establishing the duality in nonlinear programming in the complex space [6]. For proof of the theorem we need the following lemmas.

LEMMA 1. *If* z, w *are any complex vectors of the same dimensions and* C *is a hermitian positive semi-definite matrix of the appropriate order, then*

$$\mathrm{Re}(z^*Cw) \leqq (z^*Cz)^{1/2}(w^*Cw)^{1/2}.$$

Proof. Apply the Cauchy-Schwarz inequality to the semidefinite inner product $[z, w] = z^*Cw$:

$$\mathrm{Re}[z, w] \leqq |[z, w]| \leqq [z, z]^{1/2}[w, w]^{1/2}.$$

LEMMA 2. *The set of all vectors* $w = A^*\xi - C\mathfrak{n}$, *where*

$$|\arg \xi| \leqq (\pi/2) - \alpha, \qquad |\arg A\mathfrak{n}| \leqq \alpha, \qquad \mathfrak{n}^* C\mathfrak{n} \leqq 1,$$

is closed.

Proof. Suppose that u is in the closure of the set. There exist sequences $\{\xi_k\}$, $\{w_k\}$, and $\{\mathfrak{n}_k\}$ such that

$$|\arg \xi_k| \leqq (\pi/2) - \alpha, \tag{5}$$

$$|\arg A\mathfrak{n}_k| \leqq \alpha, \tag{6}$$

$$\overset{*}{\mathfrak{n}}_k C\mathfrak{n}_k \leqq 1 \tag{7}$$

$$A^*\xi_k - C\mathfrak{n}_k = w_k \tag{8}$$

and $\{w_k\}$ converges to u.

Let us assume temporarily that the sequence $\{\mathfrak{n}_k\}$ has a limit point $\mathfrak{n}$. It follows from (6) that

$$|\arg A\mathfrak{n}| \leqq \alpha \tag{9}$$

and the inequality (7) yields

$$\mathfrak{n}^*C\mathfrak{n} \leqq 1. \tag{10}$$

Now for each complex vector z satisfying

$$|\arg Az| \leqq \alpha \tag{11}$$

it follows from (5) that

$$\operatorname{Re}(\xi_k^* A z) \geqq 0. \tag{12}$$

As $w_k \to u$ and $\mathfrak{n}_k \to \mathfrak{n}$, therefore (8) and (12) yield

$$\operatorname{Re}\{(u^* + \mathfrak{n}^* C)z\} \geqq 0. \tag{13}$$

Using Levinson's generalization of Farkas' theorem for a complex space [**4**], we see from (12) and (13) that there exists a ξ such that

$$A^*\xi = u + C\mathfrak{n}, \qquad |\arg \xi| \leqq (\pi/2) - \alpha,$$

which shows that u belongs in the set w.

Now we shall show that the sequence $\{\mathfrak{n}_k\}$ can always be chosen so that it has a limit point $\mathfrak{n}$. Let $\{\delta_k\}$ be a sequence with elements δ_k in the complex n-space and $H = \{R \mid R$ is a nonnegative integer and there exists a k_0 such that for all $k \geqq k_0$, δ_k has at most R components with real part non-zero$\}$. Then H has a least element—call this r. We now define the *(pseudo) norm* of $\{\delta_k\}$ to be the least element of H, say

$$|\{\delta_k\}| = r.$$

Since

$$|\{A\mathfrak{n}_k\}| + |\{\mathfrak{n}_k\}| \leqq m + n \tag{14}$$

it is always possible to choose $\{\mathfrak{n}_k\}$ and $\{\xi_k\}$ satisfying (5), (6) and (7) such that (14) is minimal. We shall show that the sequence $\{\mathfrak{n}_k\}$ indeed has a limit point in that case.

Assume that the sequence $\{\mathfrak{n}_k\}$ has no limit point, so that

$$|\mathfrak{n}_k| = (\mathfrak{n}_k\mathfrak{n}_k^*)^{1/2} \to \infty.$$

We may assume that $|\mathfrak{n}_k| > 0$ for all k. Let $z_k = \mathfrak{n}_k|\mathfrak{n}_k|^{-1}$. Then $\{z_k\}$ is bounded; hence there is a complex vector z such that some subsequence of $\{z_k\}$ converges to z, where $|z| = 1$, $|\arg Az_k| \leqq \alpha$, and $z_k^* C z_k \leqq 1/|\mathfrak{n}_k|^2$ for all k. Thus it follows that

$$|\arg Az| \leqq \alpha \tag{15}$$

and

$$z^* C z = 0. \tag{16}$$

As a consequence of (16), Lemma 1 yields

$$\operatorname{Re}(w^* C z) = 0 \tag{17}$$

for each complex vector w. Also the complex vector z is the limit of z_k. If z has a component whose real part is non-zero, then all but finitely many $\mathfrak{n}_k$ must have the same component with real part non-zero. Hence if λ_k is any sequence of real numbers, then

$$|\{\mathfrak{n}_k - \lambda_k z\}| \leqq |\{\mathfrak{n}_k\}|.$$

Assume first that the inequality (15) is true in the form $\mathrm{Re}(Az) \neq 0$. Let A_j, $j=1, 2, \cdots, m$, denote the jth row of A, and let

$$\lambda_k = \min\left\{\frac{\mathrm{Re}(A_j\mathfrak{n}_k)}{\mathrm{Re}(A_jz)}\middle| j = 1, 2, \cdots, m, \text{ and } \mathrm{Re}(A_jz) > 0\right\}.$$

Clearly

$$\mathrm{Re}\{A(\mathfrak{n}_k - \lambda_k z)\} \geqq 0.$$

On the other hand the relation (17) implies that

$$(\mathfrak{n}_k - \lambda_k z)^* C(\mathfrak{n}_k - \lambda_k z) = \overset{*}{\mathfrak{n}}_k C\mathfrak{n}_k \leqq 1.$$

This demonstrates the existence of the sequence $\{\mathfrak{n}_k - \lambda_k z\}$ satisfying (6) and (7). Again the manner in which λ_k is selected shows that for sufficiently large k, eventually each $A(\mathfrak{n}_k - \lambda_k z)$ has at least one more component whose real part is zero than $A\mathfrak{n}_k$, thus contradicting the minimality assumption in (14). Hence $\mathrm{Re}(Az) = 0$.

Since $z \neq 0$, we can define λ_k so that for sufficiently large k, $\mathfrak{n}_k - \lambda_k z$ has at least one more component with real part zero than $\mathfrak{n}_k$. Therefore

$$|\{\mathfrak{n}_k - \lambda_k z\}| < |\{\mathfrak{n}_k\}|, \quad \text{and} \quad |\{A(\mathfrak{n}_k - \lambda_k z)\}| = |\{A\mathfrak{n}_k\}|.$$

Hence

$$|\{\mathfrak{n}_k - \lambda_k z\}| + |\{A(\mathfrak{n}_k - \lambda_k z)\}| < |\{\mathfrak{n}_k\}| + |\{A\mathfrak{n}_k\}|,$$

again contradicting the minimality of (14). This completes the proof.

Proof of the Theorem. To prove sufficiency, assume that there exists vectors $\mathfrak{n}$ and ξ satisfying (1) and (3) respectively such that

$$\mathfrak{n}^* C\mathfrak{n} \leqq 1, \qquad A^*\xi = b + C\mathfrak{n}.$$

From (2) follows the relation

$$z^* A^* \xi = z^* b + z^* C\mathfrak{n}. \tag{18}$$

But $|\arg\{(Az)_j\xi_j\}| \leqq |\arg(Az)_j| + |\arg \xi_j| \leqq \pi/2$ using (1) and (3). Thus $\mathrm{Re}((Az)^*\xi) \geqq 0$. Hence (18) yields

$$\mathrm{Re}(z^* b + z^* C\mathfrak{n}) \geqq 0. \tag{19}$$

This last inequality implies by (4) and Lemma 1 that

$$\mathrm{Re}(b^* z) + (z^* C z)^{1/2} \geqq 0.$$

To prove the necessity, we assume that for each vector z such that $|\arg Az| \leqq \alpha$ it is true that

$$\mathrm{Re}(b^* z) + (z^* C z)^{1/2} \geqq 0.$$

We now suppose that

$$A^*\xi - C\mathfrak{n} = w \neq b \tag{20}$$

for any ξ and $\mathfrak{n}$ satisfying (1), (3), and (4). In particular we can take $\mathfrak{n}=\mathbf{0}$ and $\xi=\mathfrak{n}=\mathbf{0}$ obtaining respectively

$$A^*\xi = w \neq b \tag{21}$$

and

$$\mathbf{0} = w \neq b. \tag{22}$$

Our supposition implies [5] that there exists $\mathfrak{n}_0$ and a scalar k such that

$$\operatorname{Re}(w^* \mathfrak{n}_0) \geqq k > \operatorname{Re}(b^* \mathfrak{n}_0) \tag{23}$$

for all w defined by (20). In particular the inequalities in (23) hold for all w defined by (21) and (22). From (21) and (23) it follows that $\operatorname{Re}(\xi^* A\mathfrak{n}_0) \geqq k$ for all ξ satisfying (3), which establishes that $\mathfrak{n}_0$ must obey the condition (1). Hence the given hypothesis together with the inequality (23) shows that

$$k > \operatorname{Re}(b^* \mathfrak{n}_0) \geqq -(\mathfrak{n}_0^* C \mathfrak{n}_0)^{1/2}. \tag{24}$$

But

$$k \leqq 0, \tag{25}$$

which follows from the inequality (23) on substitution of $w=\mathbf{0}$ from (22). Thus the relation $(\mathfrak{n}_0^* C\mathfrak{n}_0)^{1/2} > 0$ is obtained from (24) and (25). Finally we remark that we can set $\xi=\mathbf{0}$ and $\mathfrak{n}=(\mathfrak{n}_0 C\mathfrak{n}_0)^{-1/2}\mathfrak{n}_0$ to obtain $w=-(\mathfrak{n}_0^* C\mathfrak{n}_0)^{-1/2} C\mathfrak{n}_0$. With this value of w, we obtain from (23) and (24)

$$(\mathfrak{n}_0^* C \mathfrak{n}_0)^{1/2} < (\mathfrak{n}_0^* C \mathfrak{n}_0)^{1/2},$$

which is a contradiction. This completes the proof of the theorem.

The author is deeply indebted to Professor Bertrand Mond for several valuable suggestions.

References

1. E. Eisenberg, Supports of a convex function, Bull. Amer. Math. Soc., 68 (1962) 192–195.
2. J. Farkas, Über die Theorie der einfachen Ungleichungen, J. Reine Angew. Math., 124 (1902) 1–24.
3. G. B. Dantzig, Linear programming and extensions, Princeton University Press, 1963.
4. N. Levinson, Linear programming in complex space, J. Math. Anal. Appl., 14 (1966) 44–62.
5. W. Fenchel, Convex cones, sets and functions, Princeton University Press, Lecture notes, 1953.
6. Davinder Bhatia, and R. N. Kaul, Non-linear programming in complex space, J. Math. Anal. Appl., 28 (1969).

MORE GENERALIZATIONS OF A THEOREM OF PAPPUS

L. E. PURSELL, University of Missouri-Rolla

Let $\mathcal{C}_P$, a plane curve of class $C^{(1)}$, move through Euclidean 3-space in such a way that it is always perpendicular to the path $\mathcal{C}_S$ of its centroid (of arc length), $\mathcal{C}_S$ being a curve of class $C^{(3)}$ such that its curvature is positive everywhere. Let $\mathcal{C}_P \times \mathcal{C}_S$ denote the surface generated. In their paper, *Generalizations of the theorems of Pappus*, this MONTHLY, 76 (1969), 355–366, (hereafter referred to as [G]), A. W. and Gary Goodman (hereafter referred to collectively as "Goodman") assume that $\mathcal{C}_P$ is fixed to the moving $(\boldsymbol{T}, \boldsymbol{N}, \boldsymbol{B})$ frame and show that if $\mathcal{C}_S$ is a plane curve, then the Pappus formula

$$\|\mathcal{C}_P \times \mathcal{C}_S\| = |\mathcal{C}_P| \cdot |\mathcal{C}_S|$$

holds (here $|\cdot|$ denotes arc length and $\|\cdot\|$ denotes area). For the special case —$\mathcal{C}_S$ is a circle—Goodman's result reduces to the Theorem of Pappus referred to in our title. But for the special case—$\mathcal{C}_P$ is a line segment and $\mathcal{C}_S$ is a circle—Goodman found that if $\mathcal{C}_P$ makes a uniform spin around its own midpoint of total amount π as it moves around $\mathcal{C}_S$, then

$$\|\mathcal{C}_P \times \mathcal{C}_S\| > |\mathcal{C}_P| \cdot |\mathcal{C}_S|.$$

They also show for the special case—$\mathcal{C}_P$ is a line segment which always has the direction of the unit normal $\boldsymbol{N}$ of $\mathcal{C}_S$—that if $\mathcal{C}_S$ is a twisted curve (i.e. its torsion is not identically zero), then again $\|\mathcal{C}_P \times \mathcal{C}_S\| > |\mathcal{C}_P| \cdot |\mathcal{C}_S|$.

In this paper we explore further the case, $\mathcal{C}_S$ is a twisted curve, and the effect of $\mathcal{C}_P$ spinning around $\mathcal{C}_S$. Using a general formula derived by Goodman, we show:

THEOREM 1. *If the curvature of $\mathcal{C}_S$ is not too great (the appropriate upper bound is given below) and if $\mathcal{C}_P$ is not a circle and is allowed to spin about $\mathcal{C}_S$ in such a way that $\|\mathcal{C}_P \times \mathcal{C}_S\|$ is minimum, then (regardless of whether $\mathcal{C}_S$ is plane or twisted) there is a unique (modulo a constant) spin function θ which gives this minimum area and this minimum area is given by the Pappus rule*

$$\|\mathcal{C}_P \times \mathcal{C}_S\| = |\mathcal{C}_P| \cdot |\mathcal{C}_S|. \tag{1}$$

If $\mathcal{C}_P$ is a circle, then, of course, spinning $\mathcal{C}_P$ has no effect on the surface generated and this surface would be a *tube surface* whose area would be given by the Pappus rule as is well known (cf. [C, p. 275, ex. 7]).

To prove Theorem 1, consider Goodman's equation [G, eq. 49]:

$$\|\mathcal{C}_P \times \mathcal{C}_S\| = \iint [(\tau + \theta')^2(ff' + gg')^2 + (1 + \kappa f \cos\theta - \kappa g \sin\theta)^2]^{1/2} dt\,ds, \tag{2}$$
$$0 \leqq s \leqq |\mathcal{C}_S|, \qquad 0 \leqq t \leqq |\mathcal{C}_P|.$$

Here as in [G], s and t denote arc length along $\mathcal{C}_S$ and $\mathcal{C}_P$, respectively, and $\tau = \tau(s)$ and $\kappa = \kappa(s)$ are the torsion and curvature of $\mathcal{C}_S$. The curve $\mathcal{C}_P$ is given by $X = f(t)$, $Y = g(t)$, where X and Y are rectangular coordinates in the plane of

$\mathcal{C}_P$ with the centroid of $\mathcal{C}_P$ at $(X, Y)=(0, 0)$. The spin function $\theta=\theta(s)$ is given by the equations

$$(3)\qquad \begin{aligned}(1, 0)_{XY} &= -\cos\theta\boldsymbol{N} - \sin\theta\boldsymbol{B},\\ (0, 1)_{XY} &= \sin\theta\boldsymbol{N} - \cos\theta\boldsymbol{B},\end{aligned}$$

where the unit vectors $(1, 0)_{XY}$ and $(0, 1)_{XY}$ are given in X, Y-coordinates and $\boldsymbol{N}$ and $\boldsymbol{B}$ are the unit normal and binormal vectors of $\mathcal{C}_S$. The angle θ is the angle from $-\boldsymbol{B}$ to $(0, 1)_{XY}$, (cf. [G, fig. 1 and eq. (33)]).

Since the integrand on the right of (2) involves a sum of squares, one obtains from standard integral inequalities:

$$(4)\qquad \|\mathcal{C}_P \times \mathcal{C}_S\| \geqq \int\int |1 + \kappa f\cos\theta - \kappa g \sin\theta|\, dt\, ds,$$

$$0 \leqq s \leqq |\mathcal{C}_S|, \qquad 0 \leqq t \leqq |\mathcal{C}_P|.$$

If $\mathcal{C}_P$ were a circle, then $(ff'+gg')(t)=0$ for all t and the equality in (4) would hold. We are assuming $\mathcal{C}_P$ is not a circle, i.e. $(ff'+gg')(t)\not\equiv 0$. Since $ff'+gg'$ is independent of s and $\tau+\theta'$ is independent of t, the equality in (4) holds, i.e. $\|\mathcal{C}_P\times\mathcal{C}_S\|$ is minimum, if and only if $(\tau+\theta')(s)\equiv 0$.

We also assume

$$(5)\qquad \kappa(s) \leqq [g(t)\sin\theta(s) - f(t)\cos\theta(s)]^{-1}$$

for all (s, t), $0\leqq s\leqq |\mathcal{C}_S|$, $0\leqq t\leqq |\mathcal{C}_P|$. (This is the upper bound on κ referred to above.) To interpret (5) geometrically, observe that (5) is equivalent to:

$$(\text{radius of curvature}) = [\kappa(s)]^{-1} \geqq g(t)\sin\theta(s) - f(t)\cos\theta(s).$$

(Recall that we assume $\kappa(s)>0$ for all s.) Now if we denote the position vector from the centroid of $\mathcal{C}_P$ to a point $(X, Y)=(f(t), g(t))$ on $\mathcal{C}_P$ by $\boldsymbol{P}(t)$ (here we are not following the notation of [G]; $\boldsymbol{P}(t)$ corresponds to their vector $\boldsymbol{R}_2(t)$), then from (3) we have

$$\boldsymbol{P}(t)\cdot\boldsymbol{N}(s) = g(t)\sin\theta(s) - f(t)\cos\theta(s).$$

Therefore (5) is equivalent to:

$$(5')\qquad [\kappa(s)]^{-1} \geqq \boldsymbol{P}(t)\cdot\boldsymbol{N}(s).$$

Hence assumption (5) means that for each value of s the curve $\mathcal{C}_P$ does not cross the polar axis (the line through the center of curvature which is parallel to the binormal $\boldsymbol{B}$ of $\mathcal{C}_S$), i.e., no part of $\mathcal{C}_P$ ever backs up as the centroid of $\mathcal{C}_P$ moves forward along $\mathcal{C}_S$. (More precisely, $\mathcal{C}_P$ must be contained in a closed, half-plane bounded by the polar axis.) This assumption is the counterpart of the requirement in the classical Pappus Theorem that $\mathcal{C}_P$ does not cross the axis of rotation.

Assuming (5) and $(\tau+\theta')(s)=0$ for all s, equation (2) becomes

$$(6)\qquad \|\mathcal{C}_P\times\mathcal{C}_S\| = \int\int dt\, ds + \int \kappa\cos\theta \int f\, dt\, ds - \int \kappa \sin\theta \int g\, dt\, ds,$$

but $\int f\,dt = \int g\,dt = 0$ since the centroid of arc length of $\mathcal{C}_P$ is at the origin of the X, Y coordinate system. Hence (6) reduces to the Pappus formula

$$\|\mathcal{C}_P \times \mathcal{C}_S\| = \iint dt\,ds = |\mathcal{C}_P| \cdot |\mathcal{C}_S|, \tag{7}$$

establishing Theorem 1.

To interpret geometrically the minimum condition, $d\theta/ds + \tau = 0$, consider a unit vector $\boldsymbol{u}$ fixed to $\mathcal{C}_P$ and lying in the plane of $\mathcal{C}_P$. Then $\boldsymbol{u}$ is of the form

$$\boldsymbol{u} = \cos(\theta + \alpha)\cdot\boldsymbol{N} + \sin(\theta + \alpha)\cdot\boldsymbol{B}, \tag{8}$$

where α is constant. Differentiating with respect to s and applying the Frenet formulas [G, eqs. (8), (9), (10)] one obtains

$$d\boldsymbol{u}/ds = (-\sin(\theta + \alpha)\cdot\boldsymbol{N} + \cos(\theta + \alpha)\cdot\boldsymbol{B})\,(\tau + \theta') - \kappa\cos(\theta + \alpha)\cdot\boldsymbol{T}. \tag{9}$$

Hence, the projection $\pi(d\boldsymbol{u}/ds)$ of $d\boldsymbol{u}/ds$ onto the plane of $\mathcal{C}_P$ vanishes if and only if $d\theta/ds + \tau$ vanishes. Since the torsion τ is sometimes interpreted as the rate of spin of the $(\boldsymbol{T}, \boldsymbol{N}, \boldsymbol{B})$ frame as it moves along $\mathcal{C}_S$, we interpret the condition $\pi(d\boldsymbol{u}/ds) \equiv 0$ to mean: $\mathcal{C}_P$ *moves without spin* as it moves along $\mathcal{C}_S$. By $\mathcal{C}_P$ moving in "a natural manner," Goodman means that $\mathcal{C}_P$ is fixed to the $(\boldsymbol{T}, \boldsymbol{N}, \boldsymbol{B})$ frame. Hence, their "movement in a natural manner" is the same as our "movement without spin" if and only if $\tau = 0$ for all s. Therefore:

THEOREM 2. *Let a continuous smooth curve $\mathcal{C}_P$ move through Euclidean 3-space in such a way that it is always perpendicular to the path $\mathcal{C}_S$ of its centroid, $\mathcal{C}_S$ being a curve of class $C^{(3)}$ and of positive curvature everywhere. If for each point of $\mathcal{C}_S$ the curve $\mathcal{C}_P$ is contained in the closed half-plane bounded by the corresponding axis of curvature of $\mathcal{C}_S$, then the area of the surface generated will satisfy $\|\mathcal{C}_P \times \mathcal{C}_S\| \geqq |\mathcal{C}_P| \cdot |\mathcal{C}_S|$. If $\mathcal{C}_P$ is not a circle, then the Pappus rule $\|\mathcal{C}_P \times \mathcal{C}_S\| = |\mathcal{C}_P| \cdot |\mathcal{C}_S|$ holds if and only if $\mathcal{C}_P$ does not spin as it moves along $\mathcal{C}_S$.*

If $\mathcal{C}_S$ is not a plane curve, then it is generally difficult to visualize the nature of the motion of $\mathcal{C}_P$ along $\mathcal{C}_S$ determined by the differential condition $(d\theta/ds) + \tau = 0$. However, if $\mathcal{C}_S$ lies on a sphere Σ, then one can give a simple description of this motion in terms of elementary geometric concepts.

Let $\boldsymbol{R}(s)$ denote the point on $\mathcal{C}_S$ given by s, $\boldsymbol{r}(s)$ denote the vector from the point $\boldsymbol{R}(s)$ to the center of the sphere Σ, and a denote the radius of Σ. Then $\boldsymbol{r}\cdot\boldsymbol{r} = a^2$ and $\boldsymbol{r}' = -\boldsymbol{R}'$. Hence $\boldsymbol{r}\cdot\boldsymbol{T} = \boldsymbol{r}\cdot\boldsymbol{R}' = -\boldsymbol{r}\cdot\boldsymbol{r}' = 0$. Therefore, $\boldsymbol{r}$ lies in the plane of $\boldsymbol{N}$ and $\boldsymbol{B}$; and for each value of s there is a corresponding angle $\alpha(s)$ such that

$$\boldsymbol{r} = a\cos(\theta + \alpha)\boldsymbol{N} + a\sin(\theta + \alpha)\boldsymbol{B}, \tag{10}$$

(cf. eq. (9) above). Differentiating both sides of (10) with respect to s, applying the Frenet formulas, and collecting like terms, one obtains

$$\begin{aligned} \boldsymbol{0} = {} & (1 + a\kappa\cos(\theta + \alpha))\boldsymbol{T} + a(\theta' + \tau + \alpha')\sin(\theta + \alpha)\boldsymbol{N} \\ & - a(\theta' + \tau + \alpha')\cos(\theta + \alpha)\boldsymbol{B}. \end{aligned} \tag{11}$$

Hence

$$0 = \theta' + \tau + \alpha'. \tag{12}$$

Therefore $\theta' + \tau = 0$ for all s if and only if α is constant, i.e., $\mathcal{C}_P$ is fixed to $\boldsymbol{r}$. Hence

THEOREM 3. *Under the hypothesis of Theorem 2, if $\mathcal{C}_S$ lies on a sphere, then $\|\mathcal{C}_P \times \mathcal{C}_S\| = |\mathcal{C}_P| \cdot |\mathcal{C}_S|$ if and only if $\mathcal{C}_P$ is fixed to the radius vector from the centroid of $\mathcal{C}_P$ to the center of the sphere.*

For a simple application of this theorem consider the case: $\mathcal{C}_P$ is a moving line segment of length $2a$ having one end point fixed at a point in 3-space and $\mathcal{C}_S$ is the path of the midpoint of $\mathcal{C}_P$. Then $\mathcal{C}_P \times \mathcal{C}_S$ is the portion of a conical surface enclosed by a sphere Σ whose center is at the vertex of the conical surface. The line containing $\mathcal{C}_P$ is the generator and $\mathcal{C}_S$ is a directrix of the conical surface. The "base" of the cone (i.e., the intersection of the conical surface with the sphere Σ) is of length $2 \cdot |\mathcal{C}_S|$. Hence, we obtain a formula,

$$\|\mathcal{C}_P \times \mathcal{C}_S\| = \tfrac{1}{2} |\mathcal{C}_P| \cdot 2 |\mathcal{C}_S| = \tfrac{1}{2}(\text{slant height}) \cdot |\text{base}|,$$

which is an extension of the familiar formula for the area of a right circular cone. Of course one can easily derive this formula in other ways without having Theorem 3. Thinking about these "spherical cones" first lead to our discovery of Theorem 3.

For a non-spherical, twisted curve one might conjecture that the directed line segment $\Gamma(s)$ from a the point $\boldsymbol{R}(s)$ on the curve $\mathcal{C}_S$ to the center of the osculating sphere $\Sigma(s)$ would satisfy the equation of motion $\theta' + \tau = 0$, but this is not true. If we fix the XY-coordinate frame to $\Gamma(s)$, then, as we shall show, the hypothesis $\theta'(s) + \tau(s) = 0$ and $\tau(s) \neq 0$ for all s implies $\mathcal{C}_S$ is a spherical curve.

A standard formula for the osculating sphere gives:

$$\boldsymbol{\Gamma} = \rho \boldsymbol{N} + (\rho'/\tau)\boldsymbol{B}, \tag{13}$$

where ρ denotes the radius of curvature, κ^{-1}. If we denote the polar coordinates of the center of $\Sigma(s)$ in the XY-plane by (a, ϕ), then from (3) and (13) we obtain

$$\rho \boldsymbol{N} + (\rho'/\tau)\boldsymbol{B} = - a \cos(\phi + \theta)\boldsymbol{N} - a \sin(\phi + \theta)\boldsymbol{B}. \tag{14}$$

Hence

$$\rho = - a \cos(\phi + \theta) \tag{15}$$

and

$$(\rho'/\tau) = - a \sin(\phi + \theta). \tag{16}$$

Since we are assuming $\boldsymbol{\Gamma}(s)$ is fixed to the XY-coordinate frame, ϕ is constant. Differentiating (15) by s and substituting from (16) gives

$$\rho' = - a' \cos(\phi + \theta) - (\rho'/\tau)\theta'. \tag{17}$$

Applying our hypothesis, $\theta' + \tau = 0$, we obtain

$$0 = -a' \cos(\phi + \theta) \tag{18}$$

for all s. Since our hypotheses imply θ is not constant, (18) implies that a, the radius of the osculating sphere, is constant. Hence by a well-known result of differential geometry (cf. [L, exercise 1.7.8]), $\mathcal{C}_S$ is either a spherical curve or its radius of curvature ρ is constant. If ρ is constant, then (13) becomes $\mathbf{\Gamma} = \rho \boldsymbol{N}$, which means that $\mathbf{\Gamma}$ and, hence, the XY-coordinate frame is fixed to the $(\boldsymbol{T}, \boldsymbol{N}, \boldsymbol{B})$. But this means θ' and, hence, τ is identically zero. Therefore the only nonplane curves that satisfy (14) are spherical curves.

Goodman's results for the hypothesis— $\mathcal{C}_S$ is a plane curve—and our Theorem 3 can be restated:

THEOREM 4. *Under the hypothesis of Theorem* 2, *if $\mathcal{C}_S$ lies on a surface Σ and Σ is either a plane or a sphere, then $\|\mathcal{C}_P \times \mathcal{C}_S\| = |\mathcal{C}_P| \cdot |\mathcal{C}_S|$ if and only if $\mathcal{C}_P$ is fixed to the normal of Σ at the centroid of $\mathcal{C}_P$.*

Are there other surfaces Σ which satisfy Theorem 4? We leave this question to the reader as a topic for further research.

References

[C] R. Courant, Differential and Integral Calculus, vol. II, Wiley-Interscience, New York, 1936.

[G] A. W. Goodman and Gary Goodman, Generalizations of the theorems of Pappus, this MONTHLY, 76 (1969) 355–366.

[L] Detlef Laugwitz, Differential and Riemannian Geometry, Academic Press, New York, 1965.

[P] Lyle E. Pursell, Generalizations of Guldin's theorem and Frenet's formulae (B.S. dissertation under S. H. Gould and M. E. Shanks), Purdue University, 1949.

[S] Dirk J. Struik, Lectures on Classical Differential Geometry, Addison-Wesley, Reading, Mass., 1950.

A FURTHER COMMENT ON PAPPUS

HARLEY FLANDERS, Purdue University

The results of Goodman and Goodman [1] and Pursell [2] can be derived in a systematic and natural way by using moving frames and differential forms [3, 4]. We begin with the structure known as a **strip** or **ribbon.**

Let $\mathbf{x} = \mathbf{x}(s)$ define a space curve C, where $0 \leqq s \leqq |C|$, and let $\mathbf{e}_1 = \mathbf{e}_1(s)$ be its unit tangent. Suppose a unit normal vector $\mathbf{e}_2 = \mathbf{e}_2(s)$ (not necessarily *the* unit normal) is attached to each point of the curve. Let $\mathbf{e}_3(s) = \mathbf{e}_1 \times \mathbf{e}_2$ so that $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ is a moving orthonormal frame. The corresponding structure equations are

$$\begin{cases} d\mathbf{x} = \mathbf{e}_1\, ds \\ d\mathbf{e}_1 = (a\mathbf{e}_2 - b\mathbf{e}_3)\, ds \\ d\mathbf{e}_2 = (-a\mathbf{e}_1 + c\mathbf{e}_3)\, ds \\ d\mathbf{e}_3 = (b\mathbf{e}_1 - c\mathbf{e}_3)\, ds, \end{cases}$$

where a, b, and c are functions of s. If the curve is closed *and* $\mathbf{e}_2$ and $\mathbf{e}_3$ return to their initial positions, then a, b, and c are periodic of period $|C|$.

Whatever we are going to move, plane curve or plane domain, will be rigidly attached to the $\mathbf{e}_2$ and $\mathbf{e}_3$ axes in the normal plane to the curve. Note that so far this approach avoids conditions on curvature; $\mathbf{x}(s)$ may perfectly well be straight for example.

Now let D be a domain in the u, v-plane of area $|D|$. Define the generic point of a solid S in space by

$$y(s, u, v) = \mathbf{x}(s) + u\, \mathbf{e}_2(s) + v\, \mathbf{e}_3(s),$$

where $0 \leqq s \leqq |C|$ and $(u, v) \in D$.

We seek the volume V of S. Of course what we really have here is a mapping

$$y\colon C \times D \to E^3.$$

If the mapping is not injective, "volume" means "algebraic volume." (See conditions below.) Compute:

$$dy = (1 - au + bv)\, ds\, \mathbf{e}_1 + (-cv\, ds + du)\, \mathbf{e}_2 + (cu\, ds + dv)\, \mathbf{e}_3.$$

Hence the volume element is given by

$$\begin{aligned} dV &= (1 - au + bv)\, ds \wedge (-cv\, ds + du) \wedge (cu\, ds + dv) \\ &= (1 - au + bv)\, ds \wedge du \wedge dv. \end{aligned}$$

Consequently

$$\begin{aligned} V &= \iiint_{C\times D} (1 - au + bv) ds \wedge du \wedge dv \\ &= |C| \cdot |D| - \int a\, ds \iint u\, du\, dv + \int b\, ds \iint v\, du\, dv. \end{aligned}$$

This is a slight improvement of Goodman's Theorem 2 because we have not assumed $\kappa \neq 0$. However, if C has a Frenet frame ($\kappa > 0$ everywhere), then we can write

$$\begin{cases} \mathbf{e}_2 = (\cos\alpha)\, \boldsymbol{n} + (\sin\alpha)\, \boldsymbol{b} \\ \mathbf{e}_3 = -(\sin\alpha)\, \boldsymbol{n} + (\cos\alpha)\, \boldsymbol{b}. \end{cases}$$

Since $\mathbf{e}_1 = \boldsymbol{t}$, we have $d\mathbf{e}_1 = \kappa \boldsymbol{n}\, ds$; this immediately implies $a = \kappa \cos\alpha$ and $b = \kappa \sin\alpha$.

Now let C_0 be a curve in the u, v-plane described by $\boldsymbol{u} = \boldsymbol{u}(t)$, where t is the arc

length and $0 \leqq t \leqq |C_0|$. Define a surface Σ by

$$z(s, t) = x(s) + u(t)\, e_2(s) + v(t)\, e_3(s).$$

Then (with $\dot{u} = du/dt$, etc.)

$$dz = (1 - au + bv)\, ds\, e_1 + (-\, cv\, ds + \dot{u}\, dt)\, e_2 + (cu\, ds + \dot{v}\, dt) e_3.$$

If n denotes the unit normal to Σ, then the element of area dA satisfies

$$\begin{aligned} dA\, n &= \tfrac{1}{2} dz \times dz \\ &= (1 - au + bv)\, ds \wedge (-\, cv\, ds + \dot{u}\, dt)\, e_3 \\ &\quad + (-\, cv\, da + \dot{u}\, dt) \wedge (cu\, ds + \dot{v}\, dt)\, e_1 \\ &\quad + (cu\, ds + \dot{v}\, dt) \wedge (1 - au + bv)\, ds\, e_2 \\ &= [-\, c(u\dot{u} + v\dot{v})\, e_1 - (1 - au + bv)\dot{v}\, e_2 \\ &\quad + (1 - au + bv)\, \dot{u}\, e_3]\, ds \wedge dt. \end{aligned}$$

Hence $dA = [c^2(u\dot{u}+v\dot{v})^2+(1-au+bv)^2]^{1/2}\, ds \wedge dt$, since $\dot{u}^2+\dot{v}^2 = |\dot{u}|^2 = 1$.

Assume $1-au+bv>0$ at each point of $C \times C_0$. Then

$$\begin{aligned} A &= \iint_{C\times C_0} dA \geqq \iint (1 - au + bv)\, ds \wedge dt \\ &= |C| \cdot |C_0| - \int a\, ds \int u\, dt + \int b\, ds \int v\, dt, \end{aligned}$$

with equality if and only if $c(s)\,(u\dot{u}+v\dot{v}) = 0$ identically. If $u\dot{u}+v\dot{v}=0$, then $u^2+v^2=k^2$ is constant, and C_0 is an arc of a circle with center at (0, 0). If C_0 is the complete circle, the formula reduces to $A = |C| \cdot |C_0|$, a special case of a known result on the volume of tubes [5]. If C_0 is not a circle, the condition for equality is $c(s) \equiv 0$.

The condition $c(s) \equiv 0$ is a linear differential equation for the normal frame $\{e_2, e_3\}$; hence there is a unique solution subject to initial data. If C happens to be a plane curve ($b \equiv 0$), this system evidently reduces to $de_3 = 0$. Hence e_3 = constant. In that case e_2 is the normal to C and e_3 is the binormal. (Recall that the assumption $\kappa \neq 0$ is unnecessary for plane curves; there is always a unique unit normal.)

This completes the results in Goodman and Goodman, and Pursell. It remains to discuss the inequality $1-au+bv>0$. In the domain case, this will guarantee that $y: C \times D \to E^3$ is regular; in the curve case, that z is regular. Of course no local condition can make these maps injective.

Here is a sufficient condition. Set

$$M = \sup\, (u^2 + v^2)^{1/2},$$

taken over D or C_0 as the case may be. Then $au - bv \leqq (a^2+b^2)^{1/2} M$, so

$(a^2+b^2)^{1/2}M<1$ is sufficient. But $(a^2+b^2)^{1/2}=|dx/ds|=\kappa$, the curvature of C. Hence a sufficient condition is

$$M < \inf_s \rho(s).$$

where $\rho(s)$ is the radius of curvature.

References

1. A. W. Goodman and Gary Goodman, Generalizations of the theorems of Pappus, this MONTHLY 76 (1969), 355–366.
2. L. E. Pursell, More generalizations of a theorem of Pappus, this MONTHLY 77 (1970) 961–965.
3. H. Flanders, Differential Forms, Academic Press, New York, 1963, pp. 40–44, 112–116.
4. ———, Differential Forms, MAA Studies in Global Analysis and Geometry, Vol. 4, Prentice-Hall, Englewood Cliffs, N. J., 1967, 57–95.
5. H. Weyl, On the volume of tubes, Amer. J. Math. 61 (1939) 461–472.

THE SIMPLE CONTINUED FRACTION EXPANSION OF e

C. D. OLDS, San Jose State College, California

1. Introduction. Students, and others, have asked where they can find a readable account of the simple continued expansion of e, namely, the beautiful result due to Euler [1] that

$$(1)\qquad \begin{aligned} e &= 2+\frac{1}{1+}\,\frac{1}{2+}\,\frac{1}{1+}\,\frac{1}{1+}\,\frac{1}{4+}\,\frac{1}{1+}\,\frac{1}{1+}\,\frac{1}{6+}\cdots \\ &= \langle 2, 1, 2, 1, 1, 4, 1, 1, 6, 1, \cdots \rangle. \end{aligned}$$

Simple continued fractions have the form

$$(2)\qquad a_1+\cfrac{1}{a_2+\cfrac{1}{a_3+\cfrac{1}{a_4+\cdots}}},$$

where a_1 is usually a positive or negative integer (but could be zero), and where the terms $a_2, a_3, a_4, \cdots$ are positive integers. It is convenient to write (2) in the form

$$a_1+\frac{1}{a_2+}\,\frac{1}{a_3+}\,\frac{1}{a_4+}\cdots,$$

with the "+" signs after the first one lowered to indicate the "step-down" process in forming a continued fraction; or simply to represent it by the symbol

$\langle a_1, a_2, a_3, \cdots \rangle$. To understand the continued fraction part of this expository paper one need only read, say, a few pages of Chapter 7 in Niven and Zuckerman [4]; noting, in particular, that the convergents of the continued fraction (2), namely,

$$c_1 = \frac{a_1}{1} = \frac{p_1}{q_1}, \; c_2 = a_1 + \frac{1}{a_2} = \frac{a_1a_2 + 1}{a_2} = \frac{p_2}{q_2}, \cdots, \tag{3}$$

can be calculated, successively, from the equations

$$\begin{aligned} p_i &= a_i p_{i-1} + p_{i-2}, \\ q_i &= a_i q_{i-1} + q_{i-2}, \end{aligned} \tag{4}$$

for $i = 1, 2, 3, 4, \cdots$, provided the *undefined* terms which occur are assigned the values $p_{-1} = 0$, $p_0 = 1$, $q_{-1} = 1$, $q_0 = 0$.

Euler derived the expansion (1) by converting the infinite series expansion of e into a continued fraction, an effective method when it succeeds. One does not always end up with a *simple* continued fraction. See Wall [6, p. 17].

Since Euler's time, mathematicians such as Lambert, Gauss, Liouville, Hurwitz, Stieltjes, to mention only a few, established continued fractions as a field worthy of independent study, with applications to many branches of mathematics. It comes as no surprise, then, that the expansion (1), and similar ones, are special cases of later developments of the subject.

However, for historical reasons, we present Hermite's derivation of (1). The main ideas are contained in his famous paper [2] in which he gave the first proof that e is a transcendental number. Hermite needed approximations to e and its integral powers, and, as a matter of fact, those he used were not convergents to their respective continued fractions. What follows, then, is not a mere translation of what Hermite wrote, but, rather, a re-working of his ideas, with changes and additions to make a self-contained exposition starting with the integral (5), given below, and ending with (1).

2. Hermite's Method. Hermite starts with an integral

$$\int e^{-rx} f(x)\, dx, \tag{5}$$

where $r \neq 0$ is an arbitrary constant, and where $f(x)$ is a polynomial of degree $n = 2m$. Repeated integration by parts transforms (5) into the form

$$\begin{aligned} \int e^{-rx} f(x) dx &= -\frac{1}{r} e^{-rx} f(x) + \frac{1}{r} \int f^{(1)}(x) e^{-rx} dx \\ &= -\frac{1}{r} e^{-rx} f(x) - \frac{1}{r^2} e^{-rx} f^{(1)}(x) + \frac{1}{r^2} \int f^{(2)}(x) e^{-rx} dx \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{aligned} \tag{6}$$

$$= -e^{-rx}\left(\frac{1}{r}f(x) + \frac{1}{r^2}f^{(1)}(x) + \cdots + \frac{1}{r^{2m+1}}f^{(2m)}(x)\right)$$
$$= -e^{-rx}\Phi(x).$$

Hence,

$$\int_0^1 e^{-rx}f(x)dx = -e^{-rx}\Phi(x)\Big]_0^1 = \Phi(0) - e^{-r}\Phi(1). \tag{7}$$

In (7) let $f(x) = x^m(x-1)^m$, then

$$f^{(j)}(0) = f^{(j)}(1) = 0, \qquad j = 0, 1, 2, \cdots, m-1,$$

consequently, the expression $\Phi(x)$ in (6) reduces for $x=0$ and $x=1$, respectively, to

$$\Phi(0) = \sum_{j=m}^{2m}\frac{f^{(j)}(0)}{r^{j+1}}, \qquad \Phi(1) = \sum_{j=m}^{2m}\frac{f^{(j)}(1)}{r^{j+1}}. \tag{8}$$

On the other hand, Taylor's expansion about $x=0$ shows that

$$f(x) = \sum_{j=m}^{2m}\frac{f^{(j)}(0)}{j!}x^j = \sum_{j=m}^{2m}\alpha_j x^j, \tag{9}$$

where $f^{(j)}(0) = j!\alpha_j$. Since $f(x)$ is a polynomial with integral coefficients, the α_j's are integers. Similarly, writing $f(x) = [(x-1)+1]^m(x-1)^m$, Taylor's expansion about $x=1$, shows that

$$f(x) = \sum_{j=m}^{2m}\frac{f^{(j)}(1)}{j!}(x-1)^j = \sum_{j=m}^{2m}\beta_j(x-1)^j, \tag{10}$$

where $f^{(j)}(1) = j!\beta_j$, the β_j's being integers. Hence, the equations (8) take the form

$$\begin{aligned}\Phi(0) &= \sum_{j=m}^{2m}\frac{j!\alpha_j}{r^{j+1}} = \frac{m!M_m(r)}{r^{2m+1}},\\ \Phi(1) &= \sum_{j=m}^{2m}\frac{j!\beta_j}{r^{j+1}} = \frac{m!N_m(r)}{r^{2m+1}},\end{aligned} \tag{11}$$

where $M_m(r)$ and $N_m(r)$ are polynomials of degree m in r with integral coefficients. Using (11), with $f(x) = x^m(x-1)^m$, we rewrite (7) in the form

$$e^rM_m(r) - N_m(r) = \frac{r^{2m+1}e^r}{m!}J_m = V_me^r, \tag{12}$$

where

$$J_m = \int_0^1 e^{-rx}x^m(x-1)^m dx, \qquad V_m = \frac{r^{2m+1}}{m!}J_m. \tag{13}$$

Setting $m=0$, $f(x)=1$, $\Phi(x)=r^{-1}$, and so $\Phi(0)=r^{-1}$, $M_0(r)=1$, $N_0(r)=1$. Similarly, for $m=1$, $f(x)=x^2-x$, and so $M_1(r)=2-r$, $N_1(r)=2+r$.

The crucial part of Hermite's development hinges on a relationship between J_m, J_{m-1}, and J_{m-2}, $m\geqq 2$, of the form $J_m+aJ_{m-1}+bJ_{m-2}=0$. To this end, integration by parts shows that

$$J_m = \frac{m}{r}\int_0^1 e^{-rx}x^{m-1}(x-1)^m dx + \frac{m}{r}\int_0^1 e^{-rx}x^m(x-1)^{m-1}dx. \tag{14}$$

In the first integral above replace $(x-1)^m$ by $(x-1)^{m-1}(x-1)$ to obtain

$$\int_0^1 e^{-rx}x^m(x-1)^{m-1}dx = \frac{r}{2m}J_m + \frac{1}{2}J_{m-1}, \tag{15}$$

and with m replaced by $m-1$,

$$\int_0^1 e^{-rx}x^{m-1}(x-1)^{m-2} = \frac{r}{2(m-1)}J_{m-1} + \frac{1}{2}J_{m-2}. \tag{16}$$

A second integration by parts shows that

$$\int_0^1 e^{-rx}x^m(x-1)^{m-1}dx = \frac{m}{r}J_{m-1} + \frac{m-1}{r}\int_0^1 e^{-rx}x^m(x-1)^{m-2}dx. \tag{17}$$

On the other hand, since $(x-1)^{m-1}=x(x-1)^{m-2}-(x-1)^{m-2}$, it follows that

$$J_{m-1} = \int_0^1 e^{-rx}x^m(x-1)^{m-2}dx - \int_0^1 e^{-rx}x^{m-1}(x-1)^{m-2}dx. \tag{18}$$

Using (16), and then solving (18) for the first integral on the right, we get

$$\int_0^1 e^{-rx}x^m(x-1)^{m-2}dx = J_{m-1} + \frac{r}{2(m-1)}J_{m-1} + \frac{1}{2}J_{m-2}. \tag{19}$$

Now, substitute (19) into (17), and follow this by equating the right side of (15) with the new form of (17). The result, after simplifications, is

$$r^2J_m - 2m(2m-1)J_{m-1} - m(m-1)J_{m-2} = 0. \tag{20}$$

Using the relationship between V_m and J_m given in (13), we can replace (20) by

$$V_m - (4m-2)V_{m-1} - r^2V_{m-2} = 0. \tag{21}$$

Next, calculate V_m, V_{m-1}, V_{m-2} from (12) and substitute into (21) to obtain

$$\begin{aligned} &e^r[M_m(r) - (4m-2)M_{m-1}(r) - r^2M_{m-2}(r)] \\ &\qquad - [N_m(r) - (4m-2)N_{m-1}(r) - r^2N_{m-2}(r)] = 0, \end{aligned} \tag{22}$$

and since e^r is irrational this demands that

$$\begin{aligned} M_m(r) - (4m-2)M_{m-1}(r) - r^2M_{m-2}(r) &= 0, \\ N_m(r) - (4m-2)N_{m-1}(r) - r^2N_{m-2}(r) &= 0. \end{aligned} \tag{23}$$

From (11) we know that $M_m(r)$ and $N_m(r)$ are polynomials of degree m in r. Hence, if we replace r by $2/k$, where k is a positive integer, then these polynomials can be written in the form

$$M_m(2/k) = S_m/k^m, \qquad N_m(2/k) = R_m/k^m, \tag{24}$$

where S_m and R_m are integers. Next, substitute the values of M_m and N_m from (24) into (12), use (13) to get

$$e^{2/k}S_m - R_m = \frac{2^{2m+1}e^{2/k}}{k^{m+1}m!}\int_0^1 e^{-2x/k}x^m(x-1)^m dx.$$

Since $|x(x-1)| \leqq 1/4$ if $0 \leqq x \leqq 1$, and since $\int_0^1 e^{-2x/k}dx < k/2$, it follows easily that

$$\left| e^{2/k}S_m - R_m \right| < \frac{e^{2/k}}{k^m m!} \cdot \tag{25}$$

Again, using (24) the relations (23) can be replaced by

$$\begin{aligned} S_m &= (4m-2)kS_{m-1} + 4S_{m-2}, \\ R_m &= (4m-2)kR_{m-1} + 4R_{m-2}. \end{aligned} \tag{26}$$

If, for convenience, we let

$$\begin{aligned} S_m + R_m &= 2^{m+1}\, T_m, \\ S_m - R_m &= -\,2^{m+1}\, Z_m, \end{aligned} \tag{27}$$

then the recurrence relations (26) can be replaced by

$$\begin{aligned} T_m &= (2m-1)kT_{m-1} + T_{m-2}, \\ Z_m &= (2m-1)kZ_{m-1} + Z_{m-2}, \end{aligned} \tag{28}$$

where, in particular,

$$\begin{aligned} &T_0 = 1,\ T_1 = k,\ T_2 = 3k^2 + 1,\ T_3 = 15k^3 + 6k, \cdots, \\ &Z_0 = 0,\ Z_1 = 1,\ Z_2 = 3k,\ Z_3 = 15k^2 + 1, \cdots. \end{aligned}$$

Referring to equations (4), this shows that T_0/Z_0, T_1/Z_1, $T_2/Z_2, \cdots$, are the convergents to the simple continued fraction $\langle k, 3k, 5k, 7k, \cdots \rangle$. Moreover, from (27) we have

$$S_m = 2^m(T_m - Z_m),\ R_m = 2^m(T_m + Z_m),$$

which transforms the inequality (25) into the form

$$\left| (e^{2/k} - 1)T_m - (e^{2/k} + 1)Z_m \right| < \frac{e^{k/2}}{(2k)^m m!},$$

and after dividing both sides by $Z_m(e^{2/k}-1)$, and noting by the recursion formula (28) that Z_m increases as m increases, we see that

$$\left|\frac{e^{2/k}+1}{e^{2/k}-1}-\frac{T_m}{Z_m}\right| < \frac{e^{2/k}}{(2k)^m m! Z_m(e^{2/k}-1)} \to 0$$

as $m \to \infty$. We are justified, then, in writing as a simple continued fraction

$$\frac{e^{2/k}+1}{e^{2/k}-1} = k + \frac{1}{3k+}\,\frac{1}{5k+}\,\frac{1}{7k+}\,\cdots = \langle k, 3k, 5k, 7k, \ldots \rangle. \tag{29}$$

Subtracting 1 from both sides of (29) gives

$$\frac{2}{e^{2/k}-1} = (k-1) + \frac{1}{3k+}\,\frac{1}{5k+}\,\frac{1}{7k+}\,\cdots. \tag{30}$$

This can give excellent approximations to e. For example, if $k=2$, then from (30) we get

$$\frac{e-1}{2} = \cfrac{1}{1+\cfrac{1}{6}\,+\cfrac{1}{10}\,+\,\cdots}\,. \tag{31}$$

Using the 7th convergent to (31), we have $(e-1)/2 \approx 342762/398959$, which shows that $e \approx 2.718281828458 \cdots$, in error by only one unit in the last place.

From (31) we see that

$$e = 1 + \cfrac{2}{1+\cfrac{1}{6}\,+\cfrac{1}{10}\,+\cfrac{1}{14}\,+\,\cdots}\,, \tag{32}$$

and it is quite easy to transform this into a simple continued fraction by the use of two transformations. Write

$$\cfrac{2}{a+\cfrac{1}{b+\cfrac{1}{c+\cdots}}} = \cfrac{2}{a+\cfrac{1}{b+\cfrac{1}{y}}}\,.$$

Then it follows easily that if a is even,

$$\cfrac{2}{a+\cfrac{1}{b+\cfrac{1}{y}}} = \cfrac{1}{\cfrac{a}{2}+\cfrac{1}{2b+\cfrac{2}{y}}}\,,$$

and if a is odd, that

$$\frac{2}{a+\cfrac{1}{b+\cfrac{1}{y}}}=\frac{1}{\cfrac{(a-1)}{2}+\cfrac{1}{1+\cfrac{1}{1+\cfrac{2}{(b-1)+\cfrac{1}{y}}}}}.$$

Thus, the continued fraction (32) transforms easily into Euler's result:

$$e=2+\frac{1}{1}+\frac{1}{2}+\frac{1}{1}+\frac{1}{1}+\frac{1}{4}+\cdots=\langle 2,\overline{1,2n,1}\rangle_{n=1}^{\infty} \tag{33}$$

where the "bar" indicated the periodic part of the fraction.

A similar discussion would show that

$$e^2=\langle 7,\overline{3n-1,1,1,3n,6(2n+1)}\rangle_{n=1}^{\infty}; \tag{34}$$

and that for $k\geqq 3$ and odd

$$e^{2/k}=\langle \overline{1,\tfrac{1}{2}[(6n+1)k-1],6k(2n+1),\tfrac{1}{2}[(6n+5)k-1],1}\rangle_{n=1}^{\infty}. \tag{35}$$

3. Conclusion. The continued fraction (29) was published by Lambert in 1761. For historical data on continued fractions see Perron [**5**]. For the technique for proving e transcendental, see Chapter 2 and 9 in Niven [**3**].

What gave Hermite the idea to start with the integral (5)? A partial answer can be found by thumbing through his previous papers to see his great skill in handling various difficult integrals involving transcendental functions. One soon senses that he must already have had in mind all the basic techniques needed to prove that e is transcendental.

References

1. L. Euler, Introductio analysin infinitorium, Lausanne 1 (1748), Chapter 18.
2. C. Hermite, Compt. Rend. Acad. Sci. Paris, 77 (1873) 18–24, 74–79, 285–293.
3. Ivan Niven, Irrational Numbers, Wiley, New York, 1956.
4. I. Niven and H. S. Zuckerman, An Introduction to the Theory of Numbers, Wiley, New York, 1960.
5. Oskar Perron, Die Lehre von den Kettenbruchen, Teubner, Leipzig, 1929.
6. H. S. Wall, Analytic Theory of Continued Fractions, Van Nostrand, Princeton, N. J., 1948.

AN ELEMENTARY MODEL FOR SKEPTICISM

DAN E. CHRISTIE, Bowdoin College

The following presentation has a number of possible uses. It can serve as an allegory to indicate how, in applied mathematics, mathematical models are fitted successively to a real life situation. It is a source of exercises on trivial set theory and pre-orders, suitable for liberal arts students at levels K-13. It provides an instance of the vivid imagery of digraphs. It also offers a non-geometric but intuitive motivation for assigning a topology to a set.

The "real life" situation. Let us consider models for certain interactions among members of a set of mathematicians, for example, members of a research seminar. Let us suppose that mathematical conjectures and theorems are intermittently generated and announced by individuals or by clusters of members. Ignoring the main business of proof and counter-example, let us concentrate on the patterns of skepticism aroused by new claims or conjectures. Credibility is a personal matter. Some mathematicians have extraordinarily reliable hunches, while for others optimism is matched only by carelessness. On the receiving end, we are familiar with many degrees of acceptance ranging from utter gullibility to extreme caution, from faith in the infallibility of established experts (or of mathematics in print) to mistrust of any proposition not proved in detail. To bring the relevant aspects of the real life situation into focus, we turn to a specific example which will serve repeatedly as a prototype.

Prototype example. Let us suppose that the membership S of a seminar is the union of three disjoint subsets, each typifying a plausible pattern of response. First, there is a clique K of individuals k_1, k_2, k_3 who trust only each other: k_i never disbelieves k_j, but otherwise k_i is a skeptic. Second, there is a cluster $L=l_1$, l_2, l_3 led by l_1. No l_k ever disbelieves an assertion proposed by l_1, but l_k too is a doubter when l_1 is not involved. Finally, there is a residual pair M of modest members m_1, m_2 who cautiously refrain from rejecting the claims of any colleague. To illustrate by special cases, a table is given.

A claim by the faction listed below on the left	is predictably disbelieved by the members listed below on the right
k_1, k_3, l_1, l_2, m_2	none
k_1, k_3, l_2, m_2	l_1, l_3
l_2, m_1	k_1, k_2, k_3, l_1, l_3
l_1	k_1, k_2, k_3
k_1, m_1	l_1, l_2, l_3
k_3, l_3	l_1, l_2

The author did his undergraduate work at Bowdoin, his graduate work at Cambridge, and at Princeton, where he received his PhD under S. Lefschetz, returning one year as a Ford Faculty Fellow. Since, he has been on the staff at Bowdoin. He directed a number of summer and academic year institutes, including several Advanced Algebra seminars, and has been active in numerous MAA committees and section offices. His publications include *Intermediate College Mechanics* (McGraw-Hill 1952) and *Vector Mechanics* (McGraw-Hill 1964). *Editor.*

This seminar S has only eight members. Obviously the same example can be expanded by enlarging the subsets or by having several substructures such as L with various patterns of leadership. We now proceed to apply mathematical paraphernalia to the class of situations typified by this example.

A preliminary set model. To say that skepticism is a personal matter is to say that our models will be naively set-theoretic. Let S denote the set of mathematicians in question—the membership of the seminar for instance. To each subset or faction F where $F\subset S$ corresponds another subset, $\mathfrak{D}F$, of disbelievers. We shall then be concerned with a disbeliever function $\mathfrak{D}$ which maps the family of subsets of S into itself:

$$\mathfrak{D}: 2^S \to 2^S.$$

We shall wish to compare $\mathfrak{D}$ with the familiar complement function

$$\mathfrak{C}: 2^S \to 2^S,$$

defined by $\mathfrak{C}F=\{x\in S\mid x\notin F\}$. For later comparison we state a few properties of $\mathfrak{C}$:

$$\mathfrak{C}\mathfrak{C}F = F \text{ (for every } F \subset S), \tag{1}$$

$$F_1 \subset F_2 \Rightarrow \mathfrak{C}F_2 \subset \mathfrak{C}F_1, \tag{2}$$

$$\mathfrak{C}\varnothing = S, \tag{3}$$

$$\mathfrak{C}(F_1 \cup F_2) = \mathfrak{C}F_1 \cap \mathfrak{C}F_2. \tag{4}$$

Property (1) states that the function $\mathfrak{C}$ is an involution. Property (2) points out that $\mathfrak{C}$ reverses inclusions. Assuming that $\mathfrak{C}S=\varnothing$, we can regard (3) as a consequence of (1). Property (4) is one of the famous De Morgan rules.

The bare language of sets permits us to write concisely additional simple statements about the prototype example:

$$K \cup L \cup M = S, \tag{5}$$

$$K \cap L = L \cap M = M \cap K = \varnothing, \tag{6}$$

$$F \subset K \Rightarrow K \cap \mathfrak{D}F = \varnothing, \tag{7}$$

$$l_1 \in F \Rightarrow L \cap \mathfrak{D}F = \varnothing, \tag{8}$$

$$\text{for all } F \subset S,\ M \cap \mathfrak{D}F = \varnothing. \tag{9}$$

These statements are part of the example, depending on the decomposition of S into special subsets.

A *preliminary set model* $(S, \mathfrak{D})$ consists of a set S and a set function $\mathfrak{D}$ subject to the restriction

$$\mathfrak{D}F \subset \mathfrak{C}F, \quad \text{for every } F \subset S. \tag{10}$$

It is clear that the prototype example satisfies the limitation stated in (10). In general, the mere existence of a set function subject to (10) does not reveal

much. If we apply (10) to S itself, we obtain

$$\mathfrak{D}S \subset \mathfrak{C}S = \emptyset,$$

a result which makes a cheerful comment on unanimity. Note that extreme and unsubtle examples also satisfy (10). For instance, suppose $\mathfrak{D}F = \emptyset$ for all $F \subset S$. Clearly (10) is satisfied. Or suppose that for all F it happens that $\mathfrak{D}F = \mathfrak{C}F$. Again (10) is satisfied. To get closer to the dynamics of disbelief, we specify additional structure.

A dominance digraph. In our prototype example, the set function $\mathfrak{D}$ was based on simple characterizations of three subsets. These special cases and others may be handled automatically by assigning a dominance relation to S and then defining $\mathfrak{D}$ in terms of that relation. We read "x dominates y" for "$x \to y$." (This notation is used by Moon in [2].) The relation $\to$ is to be reflexive and transitive:

(11) $$x \to x \text{ for every } x \in S;$$

(12) $$\text{if } x \to y \quad \text{and} \quad y \to z, \quad \text{then} \quad x \to z.$$

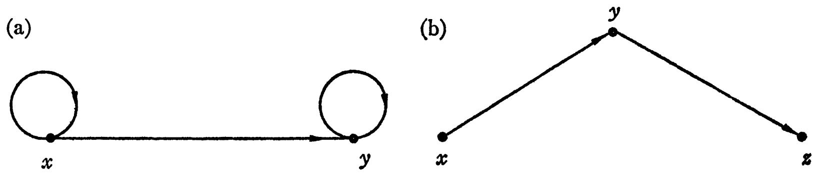


FIG. 1

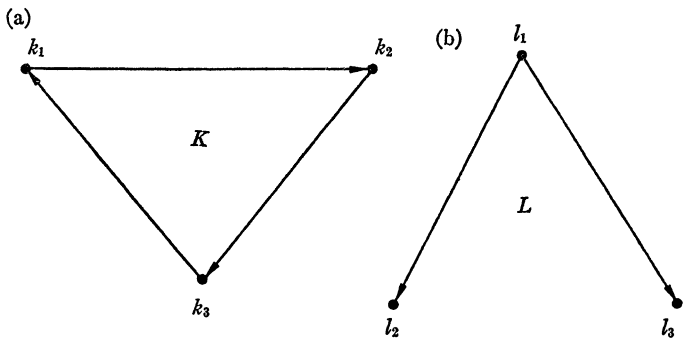


FIG. 2

The relation $\to$ is thus a pre-order. It can be conveniently portrayed by a directed linear graph (digraph), just by replacing the statement $x \to y$ by the picture in Figure 1a. (For an extensive treatment of digraphs, see [1].) The loops at the vertices, required by reflexivity, are often omitted from diagrams. Similarly, in Figure 1b three loops of reflexivity are omitted as is the arrow $x \to z$ which is dictated by transitivity.

In the prototype example the clique K can be represented as in Figure 2a. (Note that other dominances given by reflexivity and transitivity are omitted.) The cluster L led by l_1 looks like Figure 2b, expressing the dominances $l_1 \rightarrow l_i$, all i. The subset M has members which are effectively dominated by everyone. The total diagram for the prototype example looks in part like Figure 3. To complete the digraph, as before, one must use (11) and (12) wherever possible.

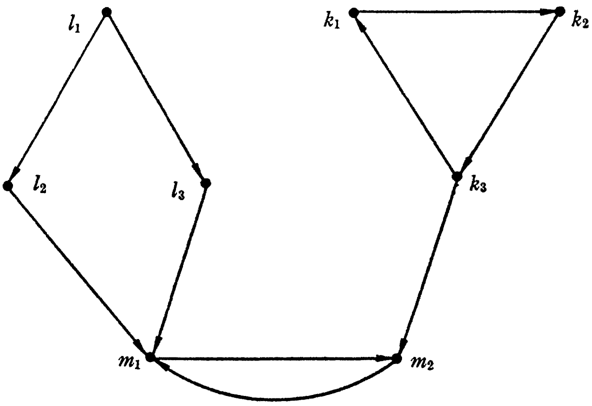


FIG. 3

Let us extend slightly the nomenclature introduced above. For a subset $F \subset S$ we shall write $F \rightarrow x$ (read "F dominates x") when there exists some $y \in F$ such that $y \rightarrow x$. We point out three simple propositions using this notation. First, a technicality based on reflexivity:

(13) $$F \not\rightarrow x \Rightarrow x \notin F.$$

And now we make two obvious remarks.

(14) $$\text{If } F_2 \not\rightarrow x \quad \text{and} \quad F_1 \subset F_2, \quad \text{then} \quad F_1 \not\rightarrow x.$$

(15) $$\text{If } F_1 \not\rightarrow x \quad \text{and} \quad F_2 \not\rightarrow x, \quad \text{then} \quad F_1 \cup F_2 \not\rightarrow x.$$

These three conclusions are immediate consequences of the definition of "$F \rightarrow x$." It is interesting to apply this extended nomenclature to the prototype example. For instance, $L \rightarrow l_j$ for all $l_j \in L$; $\{l_1\} \rightarrow l_j$ for all $l_j \in L$; $K \rightarrow m_i$ and $L \rightarrow m_i$ for all $m_i \in M$.

A dominance model. In the preliminary set model, we postulated a disbelief function $\mathfrak{D}$ subject only to (10). Now, assuming a pre-ordered set $(S, \rightarrow)$, we define $\mathfrak{D}$ explicitly: the set of disbelievers of a faction consists of all members of S which are not dominated by a member of the faction. Symbolically,

(16) $$\mathfrak{D}F = \{x \in S \mid F \not\rightarrow x\}.$$

A neater statement describes the complement of $\mathfrak{D}F$:

(17) $$x \in \mathfrak{C}\mathfrak{D}F \quad \text{iff} \quad F \rightarrow x.$$

For any pre-ordered set $(S, \rightarrow)$, we now have a dominance model $(S, \rightarrow, \mathfrak{D})$ with $\mathfrak{D}$ defined as above. A digraph illustration is given in Figure 4. The arrows of reflexivity and transitivity are omitted. Dashed lines indicate a sample F and the corresponding $\mathfrak{D}F$.

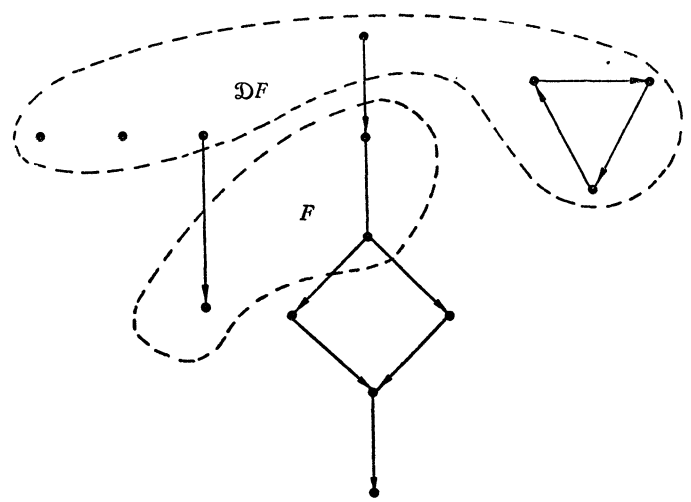


FIG. 4

It is soon clear that the dominance model $(S, \rightarrow, \mathfrak{D})$ agrees with our preliminary set model. The extreme cases cited earlier ($\mathfrak{D} = \mathfrak{C}$ and, for F not empty, $\mathfrak{D}F = \varnothing$) now appear as immediate consequences of extreme pre-orders. If $x \in \mathfrak{D}F$, then $F \not\rightarrow x$, so by (13) we know $x \notin F$. Hence (10) is verified as is its corollary, $\mathfrak{D}S = \varnothing$. For completeness, let us also verify

$$\mathfrak{D}\varnothing = S. \tag{18}$$

Obviously, $\mathfrak{D}\varnothing \subset S$. Suppose $x \in S$ but $x \notin \mathfrak{D}\varnothing$. Then by (17), we have $\varnothing \rightarrow x$ which is impossible. A more significant result is the following:

LEMMA. *In a dominance model* $(S, \rightarrow, \mathfrak{D})$

$$F_1 \subset F_2 \Rightarrow \mathfrak{D}F_2 \subset \mathfrak{D}F_1. \tag{19}$$

Proof. Let $x \in \mathfrak{D}F_2$. Then $F_2 \not\rightarrow x$ and by (14) $F_1 \not\rightarrow x$. It really is sufficient to point out that (19) is a rewording of (14).

We now confront a major consequence of our definition of $\mathfrak{D}$.

PROPOSITION. *In a dominance model* $(S, \rightarrow, \mathfrak{D})$, *for any two subsets* F_1 *and* F_2 *of* S, *it follows that*

$$\mathfrak{D}(F_1 \cup F_2) = \mathfrak{D}F_1 \cap \mathfrak{D}F_2. \tag{20}$$

Proof. The inclusions $F_1 \subset F_1 \cup F_2$ and $F_2 \subset F_1 \cup F_2$ enable us to conclude directly from (19) that $\mathfrak{D}(F_1 \cup F_2) \subset \mathfrak{D}F_1 \cap \mathfrak{D}F_2$. Now suppose that $x \in \mathfrak{D}F_1 \cap \mathfrak{D}F_2$.

Then by definition of $\mathfrak{D}$, $F_1 \not\rightarrow x$ and $F_2 \not\rightarrow x$. Hence by (15), $F_1 \cup F_2 \not\rightarrow x$; so $x \in \mathfrak{D}(F_1 \cup F_2)$.

The assertion about disbelief contained in (20) has a degree of commonsense plausibility. In general, we might expect disbelief to diminish in the face of a broader consensus, and, more specifically, the familiar De Morgan pattern (compare (4)) is perhaps reassuring. If (20) is extended, by an easy induction, to

$$\mathfrak{D}(F_1 \cup \cdots \cup F_k) = \mathfrak{D}F_1 \cap \cdots \cap \mathfrak{D}F_k \tag{21}$$

for any integer $k > 2$, we can observe that a convenient formula for calculation has emerged. Suppose that $F = \{x_1, x_2, \cdots, x_k\}$. Then

$$\mathfrak{D}F = \mathfrak{D}(\{x_1\} \cup \cdots \cup \{x_k\}) = \mathfrak{D}\{x_1\} \cap \cdots \cap \mathfrak{D}\{x_k\}. \tag{22}$$

If we can tabulate disbelievers for each individual in a faction, then the disbelievers for the faction as a whole are those who disbelieve every member of the faction. The preceding conclusion about $(S, \rightarrow, \mathfrak{D})$ gives us warning that this model is as rigid as the original set model was flexible. Yet the dual De Morgan equation does not hold. We have instead

$$\mathfrak{D}(F_1 \cap F_2) \supset \mathfrak{D}F_1 \cup \mathfrak{D}F_2. \tag{23}$$

To see that equality need not hold, let F_1 and F_2 be the singletons $\{l_1\}$ and $\{l_2\}$ (referring to Figure 3). Then $\mathfrak{D}(F_1 \cap F_2) = \mathfrak{D}\varnothing = S$, but $\mathfrak{D}F_1 \cup \mathfrak{D}F_2 = K \cup \{l_1\}$. The proof of (23) is immediate from (19), using the inclusions $F_1 \cap F_2 \subset F_1$ and $F_1 \cap F_2 \subset F_2$.

Note that, with one exception, analogues for $\mathfrak{D}$ have been found of the four displayed properties for $\mathfrak{C}$: (19) corresponds to (2), (18) to (3), and (20) to (4). We now look for an analogue of (1). We rule out the obvious candidate, $\mathfrak{D}\mathfrak{D}F = F$, because it can be combined with (20) to establish an equality in place of (23). An alternate form of (1), $\mathfrak{C}\mathfrak{C}\mathfrak{C}F = \mathfrak{C}F$, suggests a less restrictive analogue which we now establish.

PROPOSITION. *In a dominance model* $(S, \rightarrow, \mathfrak{D})$, *for every* $F \subset S$ *we have*

$$\mathfrak{D}\mathfrak{C}\mathfrak{D}F = \mathfrak{D}F. \tag{24}$$

(Note that parentheses have been consistently omitted. We could write $\mathfrak{D}(\mathfrak{C}(\mathfrak{D}F))$ for $\mathfrak{D}\mathfrak{C}\mathfrak{D}F$.)

Proof. First, one inclusion is immediate; for $\mathfrak{D}F \subset \mathfrak{C}F$ by (10), hence by (2) $\mathfrak{C}\mathfrak{D}F \supset \mathfrak{C}\mathfrak{C}F$. That is, by (1), $\mathfrak{C}\mathfrak{D}F \supset F$. Finally, by (19), $\mathfrak{D}\mathfrak{C}\mathfrak{D}F \subset \mathfrak{D}F$.

Suppose now that the opposite inclusion does not hold, i.e., for some particular F and some y in $\mathfrak{D}F$ we have $y \notin \mathfrak{D}\mathfrak{C}\mathfrak{D}F$, i.e., $y \in \mathfrak{C}\mathfrak{D}\mathfrak{C}\mathfrak{D}F$. By (17) then $\mathfrak{C}\mathfrak{D}F \rightarrow y$ which means by definition of $\rightarrow$ that there is some $z \in \mathfrak{C}\mathfrak{D}F$ such that $z \rightarrow y$. Again by (17) $F \rightarrow z$. Hence by transitivity $F \rightarrow y$, contrary to the assumption that $y \in \mathfrak{D}F$.

An unsurprising consequence of (24) is the following

COROLLARY. *In a dominance model* $(S, \rightarrow, \mathfrak{D})$, *if* $\mathfrak{D}\mathfrak{D}F$ *is equal to* F *for every faction* $F \subset S$, *then* $\mathfrak{D}$ *is equal to the complement function* $\mathfrak{C}$.

Conclusion (24), perhaps more than (22), may make the model $(S, \rightarrow, \mathfrak{D})$ appear untenable. Do we indeed expect the disbelievers of the nondisbelievers of a faction F to be indistinguishable from the disbelievers of F itself? To test the model further, let us designate as *subservient* any faction F for which $\mathfrak{D}\mathfrak{D}F$ is empty. It is easy now to show that if S has a subservient faction F, then $\mathfrak{D}F$ dominates every member of S. In the next section, we suggest an approach which allows us to retain or discard (24) as we please.

Axiomatic models. Our preliminary set model $(S, \mathfrak{D})$ required too little of $\mathfrak{D}$. The dominance model $(S, \rightarrow, \mathfrak{D})$ may have required too much, but at least it provided several properties which we may or may not wish to retain.

(10) $$\mathfrak{D}F \subset \mathfrak{C}F \qquad \text{I}$$

(18) $$\mathfrak{D}\varnothing = S \qquad \text{II}$$

(20) $$\mathfrak{D}(F_1 \cup F_2) = \mathfrak{D}F_1 \cap \mathfrak{D}F_2 \qquad \text{III}$$

(24) $$\mathfrak{D}\mathfrak{C}\mathfrak{D}F = \mathfrak{D}F. \qquad \text{IV}$$

Another candidate for special status, Lemma (19), is an immediate consequence of III, using the equivalence of $F_1 \subset F_2$ with $F_1 \cup F_2 = F_2$ and $F_1 \cap F_2 = F_1$. One is tempted to seek consequences of I, II, and III, omitting IV, or substituting some other restriction. But it is well to point out that we are working near familiar ground. Let us apply the operator $\mathfrak{C}$ to both sides of I, II, III, and IV. We get

$$\mathfrak{C}\mathfrak{D}F \supset \mathfrak{C}\mathfrak{C}F = F \qquad \text{I}'$$

$$\mathfrak{C}\mathfrak{D}\varnothing = \mathfrak{C}S = \varnothing \qquad \text{II}'$$

$$\mathfrak{C}\mathfrak{D}(F_1 \cup F_2) = \mathfrak{C}(\mathfrak{D}F_1 \cap \mathfrak{D}F_2) = \mathfrak{C}\mathfrak{D}F_1 \cup \mathfrak{C}\mathfrak{D}F_2 \qquad \text{III}'$$

$$\mathfrak{C}\mathfrak{D}\mathfrak{C}\mathfrak{D}F = \mathfrak{C}\mathfrak{D}F. \qquad \text{IV}'$$

Regarding the combined operator $\mathfrak{C}\mathfrak{D}$ as a closure operator, we recognize in I′, II′, III′, and IV′ the Kuratowski closure axioms for a topological space. In fact, axioms I, II, III, and IV are also axioms for a topological space in terms of an exterior operator $\mathfrak{D}$. The connections between orderings and topologies are well known; e.g., see Sharp [3]. Students of dominance, skepticism, hostility, or other personal traits easily caricatured by a pre-order relation may find the world of topological spaces at least loosely related to their own milieus.

Conclusions. At the beginning it was remarked that this presentation could be regarded, among other things, as a source of exercises on sets, digraphs, and finite topologies. Students can be invited to set up specific domination models based on their own or imagined social situations. Digraphs and corresponding connectivity matrices are a source of further explorations. Finally, if one calls a subset $F \subset S$ *open* when $F = \mathfrak{D}M$ for some subset $M \subset S$, one can proceed to

verify the standard open-set axioms for a topological space and to check these axioms with specific examples. The closure operator $\mathfrak{C}\mathfrak{D}$ and the interior operator $\mathfrak{D}\mathfrak{C}$ are, of course, as susceptible to mathematical whimsy as the exterior operator $\mathfrak{D}$.

The author is indebted to his colleague, C. A. Grobe, Jr., for reading a preliminary draft and for pointing out infelicities, and to the Department of Mathematical Sciences at New Mexico State University for secretarial assistance during preparation of the final manuscript.

The slight mathematical content of this article was first used by the author as the core of his inaugural lecture "A Biased Sample of One" as Wing Professor of Mathematics at Bowdoin College on March 21, 1966. Completely rearranged, it served again for an address "Postulates About People: An Introduction to Finite Topology" at the Annual Meeting of NCTM in Philadelphia on April 19, 1968. A revised version was used at the Boston Meeting of NCTM on Nov. 14, 1968, after a "dry run" under the title "The Topology of Antagonism" at a Bowdoin mathematics major meeting on October 1.

References

1. F. Harary, R. Z. Norman, and D. Cartwright, Structural Models: An Introduction to the Theory of Directed Graphs, Wiley, New York, 1965.

2. J. W. Moon, Topics on Tournaments, Holt, Rinehart and Winston, New York, 1968.

3. H. Sharp, Jr., Quasi-orderings and topologies on finite sets, Proc., Amer. Math. Soc., 17 (1966) 1344–1349.

A CORRECTION

EDWIN E. MOISE, Harvard University

On page 454 of the current volume of the MONTHLY, it was stated that the MAA was a sponsor of the Chauvenet Symposium at the United States Naval Academy. This is not correct. The meeting was sponsored by the local Section, and replaced the regular Sectional meeting. But the MAA has never sponsored one of its own Section meetings, and does not sponsor other local meetings unless it has participated in their planning or in their financial support.

MATHEMATICAL NOTES

EDITED BY DAVID DRASIN

Manuscripts for this Department should be sent to David Drasin, Division of Mathematical Sciences, Purdue University, Lafayette, IN 47907.

A CONVERSE OF THE STONE-WEIERSTRASS THEOREM

R. E. MULLINS, Marquette University

A familiar formulation of the Stone-Weierstrass theorem states that *a subalgebra A of $C(X)$ is dense in $C(X)$ if A contains constants and separates points*. Here $C(X)$ denotes the sup norm algebra of all real-valued continuous

functions on a compact Hausdorff space X, that is, $\|f\| = \sup_{x \in X} |f(x)|$. The purpose of this short note is to show that the Stone-Weierstrass theorem can be stated in a setting which makes a converse meaningful, and to prove that the converse is true.

For any set X, the notation $B(X)$ will denote the algebra of all bounded real-valued functions with sup norm defined on X. A subalgebra A of $B(X)$ will be called an *F-algebra* if A is closed relative to uniform convergence, contains constants, and separates the points of X. A subalgebra A of $B(X)$ will be called a ***minimal algebra in*** $B(X)$ if A is an F-algebra which properly contains no other F-algebra. The Stone-Weierstrass theorem then states that $C(X)$ is a minimal algebra in $B(X)$ if X is endowed with some compact Hausdorff topology. The following theorem is a converse of the Stone-Weierstrass theorem:

THEOREM. *If A is a minimal algebra in $B(X)$, then $A = C(X)$ for some compact Hausdorff topology on X.*

Proof: Let S denote the space of all nonzero real homomorphisms on A. For each f in A, let $\hat{f}$ be the function defined on S as follows: $\hat{f}(\phi) = \phi(f)$ for each ϕ in S. If $\hat{A} = \{\hat{f} : f \in A\}$, and S is given the weakest topology that makes all functions in $\hat{A}$ continuous on S, it is well known that this topology is compact Hausdorff and $\hat{A} = C(S)$. (See [1], Theorem 20, p. 276.) Each point of X may be identified with an element of S, i.e., x is identified with ϕ_x, where $\phi_x(f) = f(x)$ for each f in A. Suppose that under this identification X is properly contained in S and $\phi \in S - X$. Let p be any element of X and set $A_p = \{f \in A : f(p) = \phi(f)\}$. Then A_p is a closed subalgebra of A containing all constants. To see that A_p is an F-algebra, let x and y be any two distinct points of X. Since $\hat{A} = C(S)$, it follows by Urysohn's lemma that there exists a function g in A such that $\hat{g}(p) = \hat{g}(\phi)$ and $\hat{g}(x) \neq \hat{g}(y)$. Observe that $g \in A_p$ and hence A_p separates points. Since A is a minimal algebra in $B(X)$, we have $A = A_p$ and $\phi = \phi_p$. This contradicts the assumption that $\phi \in S - X$. We conclude that $X = S$ and $A = C(X)$.

I wish to thank Steve Friedberg for suggesting a shortening of my original proof.

Reference

1. N. Dunford and J. Schwarz, Linear Operators, Part 1, Interscience Publishers, New York.

FROBENIUS GROUPS AND WEDDERBURN'S THEOREM

R. P. BURN AND D. M. MADURAM, St. John's College, Tamilnadu, India

Ebey and Sitaram [1] have appealed to the geometry of a Desarguesian plane in constructing a new proof of Wedderburn's theorem that every finite division ring is commutative. Their technique was to construct a Desarguesian plane from a given finite division ring k, and to consider a dilatation group of this plane. Such a dilatation group is for geometric reasons a Frobenius group in which a Frobenius complement is isomorphic to the multiplicative group k^*.

The remainder of the proof is a discussion of the structure of Frobenius complements.

It is, however, simpler to construct a suitable Frobenius group from the division ring k, by considering the group of transformations of the form

$$x \to xa + b, \qquad a \neq 0,$$

of k onto itself. This is a doubly transitive Frobenius group in which a Frobenius complement (the subset of transformations of the form $x \to xa$, $a \neq 0$) is isomorphic to k^*. Ebey and Sitaram's discussion of the structure of Frobenius complements then completes the proof of Wedderburn's theorem.

It is true that this Frobenius group has a geometric representation as the subgroup of a dilatation group of a Desarguesian plane which fixes a line other than the axis, but we have now shown that the geometrical argument is an excursion which it is possible to short-cut rather than an integral part of their proof.

Reference

1. S. Ebey and K. Sitaram, Frobenius groups and Wedderburn's theorem, this MONTHLY, 76 (1969), 526–528.

A NONSTANDARD EXAMPLE OF A DISTRIBUTION

LEE PENG-YEE, University of Auckland, New Zealand

It is known ([1] p. 74) that every locally Lebesgue integrable function defines a distribution in the sense of Laurent Schwartz. There are other functions, however, that are not locally Lebesgue integrable but which still generate distributions, for example, the derivative of the function

$$F(0) = 0, \qquad F(x) = x^2 \sin 1/x^2, \qquad x \neq 0. \tag{1}$$

Therefore it is not without interest to consider the above result for more general integrals, such as the Perron integral ([2] p. 312). In this note, we shall present a simplified version first, and then indicate how the general case for the Perron integral may be obtained.

If a proposition $p(x)$ is true for every x in $[a, b]$ except perhaps a countable number of points, we say that $p(x)$ is true **nearly everywhere** on $[a, b]$. A real-valued function f is said to be **J-integrable** on $[a, b]$ if there is a continuous function F whose derivative exists nearly everywhere and is equal to f nearly everywhere on $[a, b]$. We may therefore define the J-integral of f on $[a, b]$ as follows:

$$\int_a^b f(x)dx = F(b) - F(a).$$

The uniqueness of the integral follows from the following

THEOREM 1. *Let F be differentiable nearly everywhere and continuous on $[a, b]$. If $F'(x) \geqq 0$ nearly everywhere on $[a, b]$, then F is nondecreasing on $[a, b]$.*

Theorem 1 is well known and is usually proved by *reductio ad absurdum*; see, for example, [2] p. 200, Theorem 34.1. Here we give a constructive proof which appears to be new.

Proof of Theorem 1. Let S be the set of points at each of which the derivative F' does not exist or if it does $F'(x)<0$. By assumption, S is countable. Let us denote S by $\{x_1, x_2, \cdots\}$. Now we shall construct an open covering of $[a, b]$. Let ε be arbitrary and $x\in[a, b]$. If $x\notin S$, i.e., $F'(x)\geqq 0$, then we can choose $\delta(x)>0$ such that for $x-\delta<u<x<v<x+\delta$

$$\frac{F(v)-F(x)}{v-x}>-\frac{\varepsilon}{2(b-a)}, \qquad \frac{F(x)-F(u)}{x-u}>-\frac{\varepsilon}{2(b-a)},$$

so that

$$\begin{aligned} F(v)-F(u) &= \{F(v)-F(x)\}+\{F(x)-F(u)\} \\ &> -\frac{\varepsilon(v-u)}{2(b-a)}. \end{aligned} \tag{2}$$

Since F is continuous at $x=x_n$, $n=1, 2, \cdots$, we can choose $\delta(x_n)>0$ such that for $x_n-\delta(x_n)<u<x_n<v<x_n+\delta(x_n)$

$$|F(v)-F(u)| < \varepsilon 2^{-n-1}. \tag{3}$$

Thus the family of all open intervals $(x-\delta(x), x+\delta(x))$ forms an open covering of $[a, b]$. It follows from the Heine-Borel covering theorem that there exists a finite subcovering and hence a division

$$a=y_0<y_1<\cdots<y_n=b$$

such that $F(y_i)-F(y_{i-1})$ satisfies either (2) or (3). Consequently,

$$F(b)-F(a)=\sum_{i=1}^{n}\{F(y_i)-F(y_{i-1})\}>-\varepsilon.$$

Since ε is arbitrary, $F(b)-F(a)\geqq 0$. Similarly, we can prove that $F(v)-F(u)\geqq 0$ whenever $a\leqq u<v\leqq b$, i.e., F is nondecreasing.

As a corollary of Theorem 1, we have

THEOREM 2. *Let f be J-integrable on $[a, b]$. If $|f(x)|\leqq M$ for $a\leqq x\leqq b$, then*

$$\left|\int_a^b f(x)dx\right| \leqq M(b-a).$$

For proof, we replace F in Theorem 1 by $M(x-a)\pm\int_a^x f(u)du$.

We remark that the J-integral is not included in the Lebesgue integral. For example, the derivative of (1) is J-integrable on $[0, 1]$ but not Lebesgue integrable since the function F is not absolutely continuous on $[0, 1]$, or even of bounded variation. It is interesting to note that the Dirichlet function:

$$f(x) = \begin{cases} 1 & \text{when } x \text{ is rational,} \\ 0 & \text{when } x \text{ is irrational,} \end{cases}$$

is also J-integrable.

In what follows, we shall show that every locally J-integrable function, i.e., a function J-integrable on any closed bounded interval on the real line, defines a distribution in the sense of Laurent Schwartz [1]. Let $\mathfrak{D}$ be the space of all complex functions on $(-\infty, \infty)$ which are indefinitely differentiable and have bounded supports. A sequence $\{\varphi_n\}$ of functions in $\mathfrak{D}$ with supports contained in the same bounded set is said to converge to a function φ of $\mathfrak{D}$ as $n \to \infty$ if the derivatives of any given order m of the φ_n converge uniformly as $n \to \infty$ to the corresponding derivative of φ. A linear functional T on $\mathfrak{D}$ is continuous if $\varphi_n \to \varphi$ in the above sense implies that the complex numbers $T(\varphi_n)$ converge to the complex number $T(\varphi)$ as $n \to \infty$. A **distribution** is a continuous linear functional on the vector space $\mathfrak{D}$. A complex function is J-integrable if each of its real and imaginary parts is.

THEOREM 3. *If f is locally J-integrable on $(-\infty, \infty)$, then*

$$T_f(\varphi) = \int_{-\infty}^{\infty} f(x)\varphi(x)dx$$

is a distribution.

Proof. Obviously T_f defines a linear functional on $\mathfrak{D}$. We shall prove its continuity. Let $\varphi_n \to \varphi$ in $\mathfrak{D}$ with supports contained in $[-a, a]$, and let

$$F(x) = \int_{-a}^{x} f(u)du, \qquad -a \leqq x \leqq a.$$

For brevity, we write ψ_n for $\varphi_n - \varphi$. Then we have

$$[F(x)\psi_n(x)]' = f(x)\psi_n(x) + F(x)\psi_n'(x)$$

nearly everywhere on $[-a, a]$. Since F is continuous on $[-a, a]$, so is $F\psi_n'$. Hence $F\psi_n'$ is J-integrable on $[-a, a]$. Further, let

$$G(x) = F(x)\psi_n(x) - F(-a)\psi_n(-a) - \int_{-a}^{x} F(u)\psi_n'(u)du.$$

Then G is continuous and $G'(x) = f(x)\psi_n(x)$ nearly everywhere. Hence $f\psi_n$ is J-integrable. Consequently,

$$\int_{-a}^{a} f(x)\psi_n(x)dx + \int_{-a}^{a} F(x)\psi_n'(x)dx = F(a)\psi_n(a) - F(-a)\psi_n(-a).$$

Note that the right hand side vanishes for sufficiently large a.

Now let $\|F\|$ denote the maximum value of $|F(x)|$ for $x \in [-a, a]$, and similarly for $\|\psi_n'\|$. It follows that for sufficiently large a

$$
\begin{aligned}
|T_f(\varphi_n - \varphi)| &= \left| \int_{-a}^{a} f(x)\psi_n(x)dx \right| = \left| \int_{-a}^{a} F(x)\psi_n'(x)dx \right| \\
&\leqq 2a\|F\| \, \|\psi_n'\|.
\end{aligned}
$$

Here we use Theorem 2. Since $\varphi_n \to \varphi$ in $\mathfrak{D}$, hence $\|\psi_n'\| = \|\varphi_n' - \varphi'\| \to 0$ as $n \to \infty$, and

$$\lim_{n\to\infty} T_f(\varphi_n) = T_f(\varphi).$$

This completes the proof.

We remark that the conclusion of Theorem 3 is still true with the J-integral replaced by the Perron integral. In view of Theorem 65.1 [2] p. 332, the proof will go through as above. It would be of interest to know whether there are any standard classes of integrals which need not generate distributions.

References

1. Laurent Schwartz, Mathematics for the physical sciences (revised English edition), Hermann, Paris, 1966.

2. E. J. McShane, Integration, Princeton, 1944.

SIMPLE ENDOMORPHISM RINGS

R. WARE AND J. ZELMANOWITZ, University of California, Santa Barbara

It is a relatively easy chore to find the abelian groups which possess a simple endomorphism ring. As determined in [1, p. 219], such an abelian group must be, up to isomorphism, either a finite sum of cyclic groups of the same prime order or a finite direct sum of copies of the rational numbers. As a consequence, a simple endomorphism ring is necessarily artinian. The proof in [1] employs the elementary structure theory of abelian groups.

In this note we obtain an analogous description of those modules over any commutative ring which have simple endomorphism rings. As a further application of our elementary ring-theoretic methods, we determine the torsion-free modules over a commutative integral domain which have a (von Neumann) regular endomorphism ring.

Throughout, R will denote an arbitrary commutative ring with identity element, M will be a unitary R-module, and we let E be the ring of R-endomorphisms of M, acting as right operators on M.

THEOREM. *The endomorphism ring of M is simple if and only if M is isomorphic to a finite direct sum of copies of the quotient field K of R/P for some prime ideal P of R.*

When this is the case, P is equal to the annihilator ideal of M, and E is isomorphic to a full matrix ring over K.

Proof. Suppose that E is simple. Set $F=$ the center of E and $P=\{r\in R \mid rM=0\}$. Then M is a faithful R/P-module, and R/P may be identified with the subring of E consisting of the endomorphisms of M given by multiplication by elements of R/P. Making this identification we have $R/P\subseteq F\subseteq E$.

F, being the center of a simple ring with identity, is a field, and consequently R/P is an integral domain and P is a prime ideal. It follows that F contains the quotient field K of R/P. M is a faithful right E-module and hence via restriction is a K-vector space. Thus M is K-isomorphic to a direct sum of copies of K. A fortiori, this is an R/P-isomorphism of M. For the usual reason, M is then R-isomorphic to a finite direct sum of copies of K; otherwise, the set of R-endomorphisms with finite dimensional image would be a proper two-sided ideal of E.

The converse is immediate.

REMARK. Schur's Lemma states that if R is any ring and M is a simple R-module then $E=\mathrm{Hom}_R(M, M)$ is a division ring. The converse is false, even for R commutative; e.g., take R the ring of rational integers and M the field of rational numbers. The following corollary shows that for commutative rings this is essentially the only exception to the converse of Schur's Lemma.

COROLLARY. *Let P be the annihilator ideal of M in R. The following two statements are equivalent.*

I. *E is a division ring.*

II. *Either*: (1) *P is a maximal ideal and M is a simple R-module* (*and $M \cong R/P$ as an R-module, $E\cong R/P$ as a ring*); *or*: (2) *P is a nonmaximal prime ideal and M is R-isomorphic to the quotient field K of R/P* (*and $E\cong K$ as a ring*).

DEFINITION: By a *regular* ring we mean one in which for every element x there exists an element y with $x=xyx$. We note that if x is an element of a ring with identity which has this property, and x is not a zero divisor, then x is a unit: for $x(yx-1)=0$ and $(xy-1)x=0$ imply that $xy=1=yx$. (This proof can be modified to show that any regular ring which possesses a nonzero divisor necessarily has an identity element.)

THEOREM. *Let M be a torsion-free module over a commutative integral domain R. Then M has a regular endomorphism ring if and only if M is isomorphic to a direct sum of copies of the quotient field K of R.*

Proof. Again let $E=\mathrm{Hom}_R(M, M)$, and assume that E is regular. Then for any $0\neq r\in R$, let ϕ_r denote the endomorphism of M defined by $m\phi_r=rm\,(m\in M)$; ϕ_r is in the center of E. We claim that ϕ_r is a unit in E. By the remark above it suffices to show that ϕ_r is not a zero-divisor in E, a fact which is an immediate consequence of the hypothesis that M be torsion-free. A fortiori, ϕ_r is an epimorphism; i.e., $rM=M$.

As r was an arbitrary nonzero element of R, M is a torsion-free divisible module, and as such is a K-module, and hence R-isomorphic to a direct sum of copies of K.

For the converse, let K denote the quotient field of R. Note that

$\mathrm{Hom}_R(M, M) = \mathrm{Hom}_K(M, M)$, and apply the fact that the endomorphism ring of a vector space is regular [2; p. 123].

References

1. L. Fuchs, Abelian Groups, Pergamon Press, New York, 1967.
2. N. McCoy, The Theory of Rings, Macmillan, New York, 1964.

RESEARCH PROBLEMS

EDITED BY VICTOR KLEE

In this Department the Monthly presents easily stated research problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Material should be sent to Victor Klee, Department of Mathematics, University of Washington, Seattle, WA 98105.

IS A BODY SPHERICAL IF ALL ITS PROJECTIONS HAVE THE SAME I.Q.?

G. D. CHAKERIAN, University of California at Davis

The Isoperimetric Theorem asserts that of all simple closed plane curves of a given perimeter, the circle encloses the largest area. If we denote the perimeter by L and area by A, then this is equivalent to the following isoperimetric inequality:

$$\frac{4\pi A}{L^2} \leqq 1, \tag{1}$$

where equality can hold only for circles. The expression on the left hand side of (1) is the "isoperimetric quotient" of the given curve. Polya [6, p. 180] abbreviates this quotient as "I.Q." and thereby obtains the following happy restatement of the Isoperimetric Theorem: "Of all simple closed plane curves, the circle has the highest I.Q."

In what follows we restrict our attention to convex bodies (a convex body in Euclidean n-space is a compact, convex subset with nonempty interior). We ask to what extent a 3-dimensional convex body is characterized by the I.Q.'s of its orthogonal projections onto planes through the origin. For example, if all projections have I.Q. $=1$, then the body is a ball, since all projections are circular disks. The title question asks: *is the ball the only convex body whose orthogonal projections all have the same* I.Q.?

If we consider a 3-dimensional convex body K, then we have a number of analogues of the isoperimetric quotient. If V, S, and M are the volume, surface area, and total mean curvature of K, then the following analogues of (1) are

valid (see [4] for proofs and further discussion of these quantities):

(2a) $$\frac{4\pi S}{M^2} \leqq 1,$$ (2b) $$\frac{3MV}{S^2} \leqq 1.$$

From (2a) and (2b) follows the well-known analogue of (1):

(2c) $$\frac{36\pi V^2}{S^3} \leqq 1.$$

In (2a) and (2c) equality can hold if and only if K is a ball. Equality can hold in (2b) in other cases. When we refer to "the I.Q. of K in the sense of (2a)" we shall mean the quotient on the lefthand side of (2a) computed for the 3-dimensional convex body K. Similarly for I.Q. in the sense of (2b) or (2c). Thus, for example, of all 3-dimensional convex bodies, the ball has the highest I.Q. in the sense of (2a) or (2c).

Can one estimate the I.Q. of a convex body in terms of the I.Q.'s of its planar sections or projections? Croft [2] attributes the following (unsettled) question to Steinhaus. Let K be a 3-dimensional convex body, and for each plane E, let $q(E\cap K)$ be the I.Q. (in the sense of (1)) of $E\cap K$ (consider only planes E such that $E\cap K$ has positive area). *Then is it true that*:

(3) $$\frac{36\pi V^2}{S^3} \geqq \inf(q(E \cap K))^2?$$

The infimum is taken over all planes E which cut K in a set of positive area. Qualitatively, this would imply that K has high I.Q. in the sense of (2c) if all its planar sections have high I.Q. in the sense of (1).

Let $A(E\cap K)$ and $L(E\cap K)$ be the area and perimeter respectively of $E\cap K$. Then one has the following integralgeometric formulas for V, S, and M:

(4) $$V = \frac{1}{2\pi}\int A(E \cap K)dE,$$

(5) $$S = \frac{2}{\pi^2}\int L(E \cap K)dE,$$

(6) $$M = \int dE.$$

These integrals are evaluated over the set of all planes E such that $E\cap K \neq \emptyset$ using the integralgeometric density dE for planes in Euclidean 3-space. See [4] for the relevant definitions. Now, if we start with (5) and apply Schwarz' inequality for integrals, we obtain

(7) $$S^2 = \frac{4}{\pi^4}\left(\int \frac{L(E \cap K)}{\sqrt{A(E \cap K)}}\ \sqrt{A(E \cap K)}\, dE\right)^2$$

$$\leqq \frac{4}{\pi^4}\left(\int \frac{L(E \cap K)^2}{A(E \cap K)} dE\right)\left(\int A(E \cap K) dE\right).$$

Bringing in (4) and (6), we convert (7) into

$$\text{(8)} \qquad \frac{S^2}{3MV} \leqq \frac{32}{3\pi^2} \frac{\int q^{-1}(E \cap K) dE}{\int dE}.$$

From (8) we obtain

$$\text{(9)} \qquad \frac{3MV}{S^2} \geqq \frac{3\pi^2}{32} \inf(q(E \cap K)).$$

The derivation must be taken with a grain of salt in case $q^{-1}(E \cap K)$ is not integrable. (Indeed, does there exist K such that $q^{-1}(E \cap K)$ is not integrable over the set of E such that $A(E \cap K) \neq 0$?)

The inequality (9) seems in any case rather crude, but it does contain a qualitative implication to the effect that K has high I.Q. in the sense of (2b) if all its planar sections have high I.Q. in the sense of (1). *Can one replace the constant* $3\pi^2/32$ *by* 1 *in the inequality* (9)?

It is possible to get better results using orthogonal projections rather than sections. For each direction (unit vector) u, let $K(u)$ denote the orthogonal projection of K onto a plane orthogonal to the direction u. Let $A(u)$ and $L(u)$ be the area and perimeter respectively of $K(u)$. Then one has the following expressions for S and M:

$$\text{(10)} \qquad S = \frac{1}{\pi} \int A(u) du,$$

$$\text{(11)} \qquad M = \frac{1}{2\pi} \int L(u) du,$$

where the integration is in each case over the unit sphere. If we start with (11) and apply Schwarz' inequality (as in the derivation of (8)) and then apply (10), we obtain

$$\text{(12)} \qquad \frac{M^2}{4\pi S} \leqq \frac{1}{4\pi} \int \frac{L(u)^2}{4\pi A(u)} du = \frac{\int q^{-1}(K(u)) du}{\int du},$$

where $q(K(u))$ is of course the I.Q. of $K(u)$ in the sense of (1). From (12) follows

$$\text{(13)} \qquad \frac{4\pi S}{M^2} \geqq \inf q(K(u)).$$

Again we have a qualitative result: *K has high* I.Q. *in the sense of* (2a) *if all its orthogonal projections have high* I.Q.

Examination of the derivation of (13) shows that equality holds if and only if $A(u)$ and $L(u)$ are both constant. Bodies for which $A(u)$ is constant are called "bodies of constant brightness," and bodies for which $L(u)$ is constant turn out to be precisely the well-known bodies of constant width. Hence equality holds in (13) if and only if K is a body of constant width which is also of constant brightness. *If K has constant width and constant brightness, then must K be a ball?* This is true if the boundary of K is sufficiently smooth (see [**1**, p. 140]), but the general case is still unsettled. Note that an affirmative answer to the title question would provide an affirmative answer to the last question. See [**5**] for an exposition of a number of interesting related problems of this type.

Recent work of Groemer [**3**] is of interest in connection with these problems. His results imply, for example, that any 3-dimensional convex body has *some* projection with I.Q. at least 1/4 and also *some* section with I.Q. at least 1/4.

References

1. T. Bonnesen and W. Fenchel, Theorie der konvexen Körper, Springer, Berlin, 1934. Reprint by Chelsea, New York, 1948.

2. H. T. Croft, Research Problems, mimeographed notes.

3. H. Groemer, On plane sections and projections of convex sets, to appear.

4. H. Hadwiger, Altes und Neues über konvexe Körper, Birkhäuser, Basel und Stuttgart, 1955.

5. V. Klee, Is a body spherical if its HA-measurements are constant? this MONTHLY, 76(1969) 539–542.

6. G. Polya, Induction and Analogy in Mathematics, Volume I of Mathematics and Plausible Reasoning, Princeton, 1954.

CLASSROOM NOTES

EDITED BY DAVID DRASIN

Manuscripts for this Department should be sent to David Drasin, Division of Mathematical Sciences, Purdue University, Lafayette, IN 47907.

NORMALITY OF LINEAR COMBINATIONS OF NON-NORMAL RANDOM VARIABLES

B. K. KALE, University of Manitoba

1. If U and V are independent normal random variables (r.v.) then it is well known that $U+V$ is also normal. The converse proposition, which asserts that if U, V are independent, and $U+V$ is normal, then each U and V must be normal, is a well-known theorem due to Cramér [1]. If the assumption of independence of U and V is dropped, then one could have the following possibilities:

(i) U and V are marginally normal and $U+V$ is also normal.
(ii) U and V are marginally normal but $U+V$ is not normal.
(iii) U and V are marginally not normal but $U+V$ is normal.

For situation (i), consider (U, V) bivariate normal. It may be noted here that U, V, and $U+V$ normal does not imply that (U, V) is bivariate normal. Indeed, the normality of each of $(\alpha_i U+\beta_i V)$, $i=1, 2, \cdots, n$, does not imply that (U, V) is bivariate normal. This follows from the characterization of bivariate normal distribution which asserts that (U, V) is bivariate normal iff every linear function $(\alpha U+\beta V)$ is univariate normal (see Rao [2]).

The situations (ii) and (iii) are not that easy to come by. Rosenberg [3] gave an example of (ii) but not of (iii). Most of the common textbook examples of (U, V) marginally normal but not jointly normal are examples of case (ii), as remarked by Ferguson [4]. In this note we consider a class of probability density functions (p.d.f.'s) which will be useful to demonstrate both situations (ii) and (iii). This is done by constructing an example of r.v.'s X, Y which are both normal but such that no linear combination $aX+bY$, $(ab\neq 0)$ is normal.

2. For any real α, $0<\alpha<1$, consider the class of p.d.f.'s of (X, Y), given by

$$\begin{aligned} g_\alpha(X, Y) &= f(x)f(y) + \alpha^2 \frac{xy}{2\pi e}, \quad |x| < 1, \ |y| < 1, \\ &= f(x)f(y) \text{ otherwise.} \end{aligned}$$

Here

$$f(y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right)$$

is the p.d.f. of standard normal r.v. By straightforward integration it is easy to show that:

$$\int g_\alpha(x, y)\, dx = f(y), \qquad y \in R_1$$

and

$$\int g_\alpha(x, y) dy = f(x), \qquad x \in R_1.$$

Thus X and Y are both normal with mean zero and variance unity. Similarly, one can show that:

$$\text{Cov}(X, Y) = \frac{\alpha^2}{2\pi e} \cdot \frac{4}{9} = \frac{2\alpha^2}{9\pi e} \cdot$$

Let $\phi(t_1, t_2)$ be the characteristic function (ch.f.) of (X, Y). Then:

$$\phi(t_1, t_2) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \exp\{it_1x + it_2y\} f(x) f(y)\, dx\, dy$$
$$+ \frac{\alpha^2}{2\pi e}\int_{-1}^{+1}\int_{-1}^{+1} \exp\{it_1x + it_2y\}\, xy\, dx\, dy.$$

Noting that the ch.f. of the standard normal r.v. is $\exp(-t^2/2)$ we have

$$\phi(t_1, t_2) = \exp\{-\tfrac{1}{2}(t_1^2 + t_2^2)\} + (\alpha^2/2\pi e)\, \psi(t_1)\, \psi(t_2), \tag{1}$$

where

$$\begin{aligned} \psi(t) &= \int_{-1}^{+1} \exp(itu)\, u\, du \\ &= 2i\int_0^1 u \sin tu\, du \\ &= \begin{cases} (2i/t^2)\{\sin t - t\cos t\} & \text{if } t \neq 0 \\ 0 & \text{if } t = 0. \end{cases} \end{aligned} \tag{2}$$

Let $Z = aX + bY$, $ab \neq 0$, then the ch.f. of Z is given by $\phi(at, bt)$. Now

$$\phi(at, bt) = \exp\{-\frac{t^2}{2}(a^2 + b^2)\} + \frac{\alpha^2}{2\pi e}\psi(at)\,\psi(bt). \tag{3}$$

If Z were normally distributed, then noting that Z has mean zero and variance, $a^2 + b^2 + (4\alpha^2/9\pi e)(ab)$, its ch.f. must be given by

$$\eta(t) = \exp\left\{-\frac{t^2}{2}\left(a^2 + b^2 + \frac{4\alpha^2}{9\pi e}\cdot ab\right)\right\},$$

and $\phi(at, bt) = \eta(t)$ for all the values of t. It is easy to show that $\phi(a, b) \neq \eta(1)$ and thus Z cannot be a normal variable. Thus we have case (ii) of X, Y being marginally normal but no linear combination $aX + bY$, $ab \neq 0$ is normal.

It is now trivial to obtain situation (iii). Define $U = aX + bY$ and $V = aX - bY$, $ab \neq 0$; then both U and V are marginally not normal but $U + V = 2aX$ and $U - V = 2bY$ are both marginally normal, since X and Y are marginally normal.

3. The above results can be generalized in a straightforward manner to the situation involving n variables. For $0 < \alpha < 1$, consider the class of p.d.f.'s of $(X_1, X_2, \cdots, X_n)$ given by

$$g_\alpha(x_1, x_2, \cdots, x_n) = f(x_1)f(x_2)\cdots f(x_n) + \frac{\alpha^n}{(\sqrt{2\pi e})^n}\, x_1x_2\cdots x_n$$

for $|x_i| < 1$, $i = 1, 2, \cdots, n$; $g_\alpha(x_1, x_2, \cdots, x_n) = f(x_1)f(x_2)\cdots f(x_n)$ otherwise.

Then by the arguments similar to those used above, we have

(1) $X_1, X_2, \cdots, X_n$ are marginally normal and any set of $r<n$ variables are jointly normal,

(2) $\sum_{i=1}^{r} a_i x_i$, $r<n$, is normal,

(3) $\sum_{i=1}^{n} a_i x_i$ with $a_1, a_2, \cdots, a_n \neq 0$ is not normal.

References

1. H. Cramér, Random Variables and probability distributions, Cambridge Tracts in Mathematics, and Mathematical Physics Tract No. 36 (1935).

2. C. R. Rao, Linear Statistical Inference, Wiley, New York, 1965, pp. 437–438.

3. Lloyd Rosenberg, Non-normality of linear combinations of normally distributed random variables, this MONTHLY, 72 (1965) 888–890.

4. Thomas S. Ferguson, Mathematical Statistics, a Decision Theoretic Approach, Academic Press, New York, 1967, 111.

A REHABILITATION OF $(1+z/n)^n$

J. VAN YZEREN, Technological University of Eindhoven

1. There are many roads to e^z. Preference for one or another has become a matter of context or personal taste. Thus it could happen that the time-honored sequence $(1+z/n)^n$ would fall into disgrace. Its "many disadvantages" even were pointed out by an authority like G. H. Hardy [1], and in fact at first sight $(1+z/n)^n$ looks rather awkward to deal with.

However, $(1+z/n)^n$ leads simply and most directly to essential results if one uses an adequate expression for the differences.

2. Starting with real $z=x$ we put

$$1+\frac{x}{n+1}=u, \qquad 1+\frac{x}{n}=v.$$

Clearly $nv=nu+u-1$. This is used in

$$\begin{aligned} u^{n+1}-v^n &= u(u^n-v^n)+(u-1)v^n = u(u^n-v^n)+n(v-u)v^n \\ &= (u-v)(u^n+u^{n-1}v+\cdots+uv^{n-1}-nv^n) \\ &= (u-v)^2(u^{n-1}+2u^{n-2}v+\cdots+nv^{n-1}). \end{aligned}$$

Thus it follows that

$$\left(1+\frac{x}{n+1}\right)^{n+1}-\left(1+\frac{x}{n}\right)^n=\frac{x^2}{n^2(n+1)^2}\sum_{k=1}^{n} ku^{n-k}v^{k-1}. \tag{2.1}$$

Here we see that for fixed $x\neq 0$, the sequence $(1+x/n)^n$ increases with n as soon as $1+x/n=v\geqq 0$ (which implies $(n+1+x)/(n+1)=u>0$). It is also obvious that either $(1+x/n)^{n^2}$ has a limit if $x<0$ or $(1-x/n)^n$ has if $x>0$, and anyway $(1-x^2/n^2)^{n^2}$ has a positive limit $\leqq 1$. Upon taking nth roots, we see that

$(1-x^2/n^2)^n=(1+x/n)^n(1-x/n)^n$ tends to 1. Consequently $(1+x/n)^n$ and $(1-x/n)^n$ tend to reciprocal limits "e^x" and "$1/e^x$". In addition the monotonicity gives

$$\left(1+\frac{x}{n}\right)^n \leqq e^x \leqq \left(1-\frac{x}{n}\right)^{-n} \quad \text{if } -1 \leqq \frac{x}{n} < 1.$$

In the special case $n=1$ we find

$$1+x \leqq e^x \leqq 1+x+x^2/(1-x), \qquad -1 \leqq x < 1,$$

which shows that e^x is differentiable at 0.

Further relevant properties ($e^x \cdot e^y=e^{x+y}$, $De^x=e^x$, $\log x$) follow along well-known lines.

3. The complex sequence $(1+ix/n)^n$ consists of components which are monotonic (in n) and bounded. It leads to $\cos x$ and $\sin x$ in basically the same way as $(1+x/n)^n$ has led to e^x. See [2]. However, we omit this because $(1+z/n)^n$ and the notion of Cauchy-sequence have a much broader scope.

We replace x in (2.1) by z and majorize $|u^{n-k}v^{k-1}|$ as follows:

$$\left|\left(1+\frac{z}{n+1}\right)^{n-k}\left(1+\frac{z}{n}\right)^{k-1}\right| \leqq \left(1+\frac{|z|}{n}\right)^n \leqq e^{|z|}.$$

(If one prefers not to use the latter inequality, a less sharp bound will do:

$$\left(1+\frac{|z|}{n}\right)^n \leqq \left(1+\frac{|z|}{kn}\right)^{kn} \leqq \left(1-\frac{|z|}{kn}\right)^{-kn} \leqq \left(1-\frac{|z|}{k}\right)^{-k} < 2^k,$$

with natural $k>2|z|$.)

Since $\sum_{k=1}^{n} k=\frac{1}{2}n(n+1)$, we find

$$\left|\left(1+\frac{z}{n+1}\right)^{n+1}-\left(1+\frac{z}{n}\right)^n\right| \leqq \frac{|z|^2}{n^2(n+1)^2}\cdot\frac{1}{2}\,n(n+1)e^{|z|}$$
$$= \frac{1}{2}\,|z|^2e^{|z|}\left(\frac{1}{n}-\frac{1}{n+1}\right)$$

and so

$$\left|\left(1+\frac{z}{m}\right)^m-\left(1+\frac{z}{n}\right)^n\right| \leqq \frac{1}{2}\,|z|^2e^{|z|}\left(\frac{1}{n}-\frac{1}{m}\right), \qquad m > n.$$

According to Cauchy's principle of convergence $\lim_{m\to\infty}(1+z/m)^m$ exists, giving

$$\left|e^z-\left(1+\frac{z}{n}\right)^n\right| \leqq \frac{1}{2}\,|z|^2e^{|z|}\,\frac{1}{n}\,.$$

The case $n=1$ shows that e^z is differentiable at $z=0$. As the sequence most easily leads to the proof that $e^{z_1}\cdot e^{z_2}=e^{z_1+z_2}$, there remains only one important property to prove: the periodicity of e^z.

4. We try to invert $e^z = w$. The possibility of $z \neq 0$ in $e^z = 1$ being of primary interest, we first consider w such that $|w| = 1$. We write $w = u$ (**unitary**) and accordingly introduce unitary sequences, for example

$$\frac{1 + ix/2}{1 - ix/2}, \left(\frac{1 + ix/4}{1 - ix/4}\right)^2, \cdots, \left(\frac{1 + ix/2n}{1 - ix/2n}\right)^n,$$

which tends to

$$e^{(1/2)ix}/e^{-(1/2)ix} = e^{ix} = u, \qquad u\bar{u} = 1.$$

With this procedure as motivation, we now seek to solve the equation $e^{i\xi} = u$, with $u \neq -1$, $|u| = 1$. We construct a sequence $x_1, x_2, \cdots, x_k$ such that

$$u = \frac{1 + ix_1/2}{1 - ix_1/2} = \left(\frac{1 + ix_2/4}{1 - ix_2/4}\right)^2, \cdots, \frac{1 + ix_{k-1}/2^{k-1}}{1 - ix_{k-1}/2^{k-1}} = \left(\frac{1 + ix_k/2^k}{1 - ix_k/2^k}\right)^2, \cdots.$$

Let $x_1 > 0$ (otherwise take $\bar{u}$); then $x_2, x_3, \cdots$ (all >0) follow from

$$\frac{1}{x_1} = \frac{1}{x_2} - \frac{x_2}{4^2}, \cdots, \frac{1}{x_{k-1}} = \frac{1}{x_k} - \frac{x_k}{4^k}, \cdots. \tag{4.1}$$

Obviously $x_1 > x_2 > x_3 > \cdots$, and $x_k \to \xi \geqq 0$. We now show for ξ that $e^{i\xi} = u$. Let us write $2^k = 2n$. Then

$$u = \left(\frac{1 + ix_k/2n}{1 - ix_k/2n}\right)^n$$

and as for unitary a and b, it follows at once from $\left|\sum_{k=1}^{n} a^{k-1}b^{n-k}\right| \leqq n$ that $|a^n - b^n| \leqq |a - b| \cdot n$. Thus

$$\left|\left(\frac{1 + i\xi/2n}{1 - i\xi/2n}\right)^n - u\right| \leqq \frac{|i\xi/n - ix_k/n|}{|1 - i\xi/2n|\,|1 - ix_k/2n|} \cdot n < x_k - \xi,$$

hence

$$e^{i\xi} = \lim_{n\to\infty}\left(\frac{1 + i\xi/2n}{1 - i\xi/2n}\right)^n = u.$$

Now let $x_1 > x_1^* > 0$; then (4.1) gives $x_2 > x_2^*, \cdots$, so $\xi > \xi^*$ ($\xi = \xi^*$ would give $u = u^*$). Even more: from

$$\frac{x_1 - x_1^*}{x_1x_1^*} = \frac{x_2 - x_2^*}{x_2x_2^*} + \frac{x_2 - x_2^*}{4^2}, \cdots,$$

we see that $x_1 - x_1^* > x_2 - x_2^* > \cdots > \xi - \xi^* > 0$. In words: ξ is a strictly monotonic and continuous function of x_1.

The limit case $x_1 = \infty$, $x_2 = 4$, $x_3 = 8(\sqrt{2} - 1)$, $\cdots$ corresponds to $u = -1$. So, writing π for the corresponding ξ, we have $e^{i\pi} = -1$. Likewise $e^{-i\pi} = -1$ corresponds to $x_1 = -\infty$, $x_2 = -4$, $\cdots$. This π is the area of the unit circle, for

∞, 4, $8(\sqrt{2}-1)$, $\cdots$ are the areas of subsequent circumscribed 2^k-gons ($k=1, 2, 3, \cdots$). The equality $e^{i\pi}=e^{-i\pi}$ displays the periodicity of e^z.

In the same "pre-infinitesimal" way $e^z=w$ can be solved for all positive w: change ix into x (that means in (4.1): $-$into$+$). For example $e^z=2$ leads to $2=(1+\frac{1}{2}x_1)/(1-\frac{1}{2}x_1)$, so $x_1=2/3$, which already approximates log 2.

References

1. E. T. Whittaker and G. N. Watson, Modern Analysis, Cambridge Univ. Press, New York, 1946, p. 581.

2. J. van Yzeren, De exponentiële functie, Nieuw Tijdschr. Wisk. (1968/69) 89–104.

A REMARK ON CHARACTERISTIC POLYNOMIALS

JOSEF SCHMID, University of Fribourg, Switzerland

The purpose of this note is to give a simple proof of the following

PROPOSITION. *If A and B are $n\times m$ and $m\times n$ matrices respectively, over a unitary, commutative ring R, I_n and I_m the unit-matrices of orders n and m, then for the characteristic polynomials of AB and BA holds the relation*

$$t^m\det(tI_n - AB) = t^n\det(tI_m - BA).$$

COROLLARY. *The characteristic coefficients of AB and BA are equal, as far as possible. In particular, the two extreme cases are $trAB = trBA$ and a (generalized) Lagrange-identity.*

COROLLARY. *If $AB=I_n$ and $BA=I_m$, then $n=m$.*

Proof. In this case the proposition yields $t^m(t-1)^n=t^n(t-1)^m$ and this implies $n=m$ (cf. [1, 2]).

Proof of the proposition. Consider the two quadratic matrices C, D of order $n+m$ over $R[t]$:

$$C=\begin{pmatrix} tI_n & A \\ B & I_m \end{pmatrix}, \qquad D=\begin{pmatrix} I_n & 0 \\ -B & tI_m \end{pmatrix}.$$

Multiplication gives

$$CD=\begin{pmatrix} tI_n - AB & tA \\ 0 & tI_m \end{pmatrix}, \qquad DC=\begin{pmatrix} tI_n & A \\ 0 & tI_m - BA \end{pmatrix},$$

and the result now follows from $\det CD=\det DC$.

References

1. H. Flanders, Elementary divisors of AB and BA, Proc. Amer. Math. Soc., 2(1951) 871–874.

2. W. G. Leavitt, Modules over commutative rings, this MONTHLY, 71 (1964) 1112–1113.

3. C. C. MacDuffee, The theory of matrices, Ergebnisse der Math., 2(1933) 23.

4. W. M. Scott, On the characteristic roots of matrix products, this MONTHLY, 48(1941) 201–203.

5. H. S. Thurston, On the characteristic equation of products of square matrices, this Monthly, 38(1931) 322–324.

6. H. W. Turnbull and A. C. Aitken, An Introduction to the Theory of Canonical Matrices, Blackie, London, 1932, p. 181.

A COMPUTER ASSISTED APPROACH TO INTEGRAL CALCULUS VIA JORDAN CONTENT

G. J. Porter, University of Pennsylvania

An increasingly large number of schools have been experimenting with the use of digital computers in freshman calculus courses. Many of these uses are purely computational. It is my feeling that the computer can be used in a meaningful way to present the basic concepts of pure mathematics. In particular I think that through a well thought out program of computer assignments the student can be led to formulate definitions for himself. As an illustration of this idea we discuss here a computer assisted approach to integral calculus via Jordan content. This course was taught to 30 freshmen at the University of Pennsylvania during the fall 1968 semester.

Before proceeding, a few words about philosophy are appropriate. There are as many notions of what constitutes freshman calculus as there are teachers of freshman calculus. Such courses range from an introduction to real variables to an applied engineering course. As a pure mathematician I believe that before one can compute mass, volume, etc., one must arrive at a mathematical definition of these quantities. The definition need not be rigorous but it must be mathematical. Thus definitions of area such as 'extent of surface' are mathematically inadequate. My use of the computer is aimed at helping the students formulate mathematical definitions of these quantities.

After a short introduction to computing and a discussion of the real number system, the students were assigned the problem of computing the area bounded by an ellipse. This assignment made explicit the necessity of a mathematical definition of area and the idea that area is a function which assigns a number to a region of the plane. Most students either divided the ellipse into vertical strips and approximated the area of these by rectangles (the single integral) or divided the plane into a rectangular grid and computed the area of those rectangles wholly contained within the ellipse (inner content) or those which had some point in common with the region (outer content).

These programs were discussed in class and the basic concepts were abstracted into the definitions of inner and outer Jordan content in the plane. (These were defined using g.l.b. and l.u.b.) When the inner content of the region equals the outer content we call their common value the area of the region.

The next assignments were to compute volume in three and four space. The students were able to see how to modify their programs for area to these cases. The fact that four dimensional space could not be visualized caused no difficulty. (It should be noted that computing volumes in n-space by taking rectangular grids is very time consuming for n larger than three.)

Thus the definitions of Jordan content in the plane were also generalized to n-space with no apparent difficulty. At this point a mathematical study of the properties of volume (additivity on disjoint sets, invariance under translation, etc.) can be presented. The level of this presentation should be appropriate for the class.

The next stage of development is to compute the mass of a planar region with a given density function, $f(x, y)$. The students, as imagined, subdivided the region into a grid, chose a point in each rectangle, and computed $\sum f(x_i, y_i)$ Area(R_i) where R_i is a rectangle in the grid and $(x_i, y_i) \in R_i$. Once again the sum is taken either over those rectangles contained in the region or over those rectangles which intersect the region.

The students have thus arrived at the idea of approximating a function by simple functions. Again their idea generalizes to n-space with no difficulty. Class discussion was then directed at the problem of which simple functions to choose and whether this choice affected the computation. Given a bounded density function one can take g.l.b. and l.u.b. on each rectangle and using these simple functions define upper and lower integrals in a manner similar to the definition of Jordan content. When these agree onede fines $\int_S f dV^n$ to be their common value $(S \subset R^n, f: S \to R)$.

Most importantly when the integral exists the choice of the simple function doesn't matter and the naive computation $\sum f(x_i)$ Vol(R_i) should be thought of as an approximation of the integral.

Returning to the student's calculation of mass in the plane, an analysis of the computer program written will reveal that the student is computing a loop within a loop; i.e., he is first summing (say) vertical strips and then is adding these up. This leads at once to the definition of iterated integral. Generalizing this one sees that the n-space integral can be thought of as an $(n-1)$-space integral of a single integral by collapsing along some coordinate. Proceeding in this manner one can now express the n-space integral as an n-fold iterated integral. In fact this was what the students did in their computer programs. Moreover, the limits of integration on each variable appear in the programs.

Another application of this point of view is to the calculation of volumes by viewing the n-space integral as a single integral of an $(n-1)$-space integral. This is particularly easy when an analytic expression is available for the $(n-1)$-space integral. Examples of this approach are the calculations of 'volumes of revolution' and 'volumes of known cross-section'. An interesting computation is that of the volume of the unit ball in n-space. To illustrate this the following problem was assigned.

Let $B(n, r) = \{(x_1, \cdots, x_n) \in R^n \mid \sum x_i^2 \leq r^2\}$.

(a) Prove by analytic methods that $\mathrm{Vol}(B(n, r)) = r^n \mathrm{Vol}(B(n, 1))$.

(b) Using the result of (a) and starting with $\mathrm{Vol}(B(1, 1)) = 2$ write a program to compute $\mathrm{Vol}(B(n, 1))$ for $n = 2, 3, \cdots, M$.

By using this inductive procedure the student may now perform a calculation which was formerly too time consuming. The fact that these volumes have

a maximum for $n=5$ and tend to zero as n increases is an amazing discovery for most students. In addition, when the student learns analytic methods for evaluating integrals he may return to this problem and derive exact expressions for $\text{Vol}(B(n, 1))$.

With the evaluation of n-space integrals reduced to the problem of evaluating single integrals the course specialized to the study of the single integral.

It is my feeling that this approach has many advantages. Among these are:

1. The student who takes only one year of calculus is exposed to the ideas in n-space as well as on the line.
2. The integral arises more naturally in the plane than on the line.
3. The student is led to the definitions by his own intuitive ideas.
4. The student is forced to make certain hazy thoughts into precise language in order to communicate with the computer.
5. The definite integral has an existence of its own and is not confused with the antiderivative.
6. After computing approximations of integrals the student is more appreciative of analytic methods. Early in the course my students were assigned the problem of computing the volume inside the sphere, $x^2+y^2+z^2=4$, and outside the cylinder, $x^2+y^2=1$. When I did this problem analytically in class and arrived at an answer of $4\pi\sqrt{3}$ the class gasped. They had worked very hard to get an answer they could not recognize in this form.
7. When the fundamental theorem of calculus is proved, it is as amazing a theorem as it should be. The fact that the limit of some complicated sums which one has to work hard to approximate can be evaluated by means of an antiderivative is astounding and shows the strength of the analytic approach to mathematics as opposed to brute force machine computation. Too often the student misses this point because he does not understand the definite integral. In fact he may think the fundamental theorem is the definition.

In conclusion let me remark that there is a tendency on the part of some of those who use the computer in teaching calculus to use it only to compute those integrals which do not yield to analytic methods. While this technique may from time to time be useful in applications, I believe it adds nothing mathematically to the calculus course. On the other hand I think the computer can be used effectively in motivating and helping the student understand the pure mathematics.

late in life. We agreed, therefore, to cooperate in starting a mathematical "Sunday School." The details of our new enterprise are not quite clear. We shall try to limit its scope, at least in the initial stages of its development. The implications of the success of such an undertaking can be very significant.

It is perhaps consistent with the nature of our New Careers mathematics program that we should conclude our account of what we have done with it by reporting that we are about to walk through another door that has come ajar.

PROBLEMS AND SOLUTIONS

EDITED BY EMORY P. STARKE

All problems (both elementary and advanced) proposed for inclusion in this Department should be sent to E. P. Starke, 1000 Kensington Ave., Plainfield, NJ 07060. Proposers of problems are urged to enclose any solutions or information that will assist the editors. Ordinarily, problems in well-known textbooks and results in generally accessible sources are not appropriate for this Department. No solutions (except those accompanying proposals) should be sent to Professor Starke.

ELEMENTARY PROBLEMS

Solutions of Elementary Problems should be sent to Problems Group, Mathematics Department, University of Maine, Orono, ME 04473. To facilitate their consideration, solutions of Elementary Problems in this issue should be typed (with double spacing) and should be mailed before February 28, 1971. Contributors (in the United States) who desire acknowledgment of receipt of their solutions are asked to enclose self-addressed stamped postcards.

E 2259. *Proposed by J. R. Isbell, State University of New York at Buffalo*

A group in which all uth powers commute with each other and all vth powers commute with each other, u and v relatively prime, is abelian.

E 2260. *Proposed by Marlow Sholander, Case Western Reserve University*

It is given for every real $n>0$ that three times the area between $y=1$ and $y=x^n$, $0\leqq x\leqq 1$, equals the area between $y=1$ and $y=x^n$, $1\leqq x\leqq a_n$. Find $\lim_{n\to 0} a_n$.

E 2261. *Proposed by R. M. Meyer, SUNY College at Fredonia, N. Y.*

Show that for each integer m, $0\leqq m<n$, the function $L_m(p)=\sum_{j=0}^{m}\binom{n}{j}p^j(1-p)^{n-j}$ is strictly decreasing for $0<p<1$.

E 2262. *Proposed by G. J. Simmons and D. B. Rawlinson, Sandia Laboratories, Albuquerque, N. M.*

For every $k(>1)$ there is a set of k positive integers whose sum and product are equal. For $k=2, 3, 4, 6, 24$ the set is unique. Is it unique for any other k?

E 2263. *Proposed by Bernard McCabe, Bell Comm. Inc., Washington, D.C.*

Suppose an urn contains m balls, each a different color. An observer draws a ball at random, records its color, and replaces it in the urn. He repeats this procedure until some color reappears for the first time. Show that the expected number of drawings is

$$E(m) = \sum_{k=1}^{m} \frac{k(k+1)m!}{(m-k)!m^k}$$

and determine the leading term in the asymptotic expansion.

E 2264. *Proposed by Erwin Just, Bronx Community College*

If F_k is the kth Fibonacci number ($F_k=F_{k-1}+F_{k-2}$, $F_1=F_2=1$), prove the following congruence relations:

(a) $$F_{7m}(1-F_{7k+1}) \equiv F_{7k}(1-F_{7m+1}) \quad (\text{mod } 377).$$

(b) $$5(F_{k-1}^2+F_{k+1}^2) \equiv 2(-1)^{k+1} \quad [\text{mod } (F_{k-1}+F_{k+1})].$$

SOLUTIONS OF ELEMENTARY PROBLEMS

Inequalities in the Products of Powers of an Ordered Set of Numbers

E 2203 [1969, 1138]. *Proposed by M. S. Klamkin, Ford Scientific Laboratory*

It is known that if $0 \leqq x_1 \leqq x_2 \leqq \cdots \leqq x_n$, $(n \geqq 3)$, then

$$x_1^{x_2} x_2^{x_3} \cdots x_n^{x_1} \geqq x_2^{x_1} x_3^{x_2} \cdots x_1^{x_n}.$$

Are there any other nontrivial permutations $\{a_i\}$ and $\{b_i\}$ such that

$$a_1^{a_2} a_2^{a_3} \cdots a_n^{a_1} \geqq b_1^{b_2} b_2^{b_3} \cdots b_n^{b_1}?$$

Solution (*adapted*) *by G. L. Watson, University College, London, England.* For $n=3$ there is no other nontrivial permutation of the x_i of the form required. For $n=4$, there are other solutions. For one such solution, note that $x_3/x_1 \geqq 1$, $x_4/x_3 \geqq 1$, $x_3-x_2 \geqq 0$, $x_3-x_1 \geqq 0$ imply

(1) $$(x_3/x_1)^{(x_3-x_2)}(x_4/x_3)^{(x_3-x_1)} \geqq 1,$$

whence (upon multiplying both sides by $x_2^{x_4}/x_3^{x_3}$)

(2) $$x_1^{x_2} x_2^{x_4} x_4^{x_3} x_3^{x_1} \geqq x_1^{x_3} x_3^{x_2} x_2^{x_4} x_4^{x_1}.$$

For $n>4$, the possibilities increase rapidly. For example, with $n=5$,

$$(x_5/x_2)^{(x_4-x_3)}(x_3/x_1)^{(x_5-x_2)} \geqq 1 \tag{1'}$$

implies

$$x_1^{x_2}x_2^{x_3}x_3^{x_5}x_5^{x_4}x_4^{x_1} \geqq x_1^{x_5}x_5^{x_3}x_3^{x_2}x_2^{x_4}x_4^{x_1}. \tag{2'}$$

Also solved by Haig Bohigian.

A Limit Problem

E 2204 [1969, 1138]. *Proposed by A. C. Segal, University of Alabama, Birmingham, and Basil Lepp, Rust Engineering Co.*

Let $f(n)=\sum_{k=1}^{\infty} 1/k$, where k has no zeros in its n-ary expansion. Prove or disprove: $\lim_{n\to\infty}[f(n)-n\log n]=0$.

Solution by N. J. Fine, Pennsylvania State University. The result is correct. Let S be the indicated set of integers. Then

$$S=\{1,2,\cdots,n-1\}\cup\bigcup_{x\in S}\{nx+1,nx+2,\cdots,nx+n-1\}.$$

Hence

$$\begin{aligned} f(n) &= \sum_{k=1}^{n-1}\frac{1}{k}+\sum_{x\in S}\sum_{k=1}^{n-1}\frac{1}{nx+k}\\ &= \sum_{k=1}^{n-1}\frac{1}{k}+\sum_{x\in S}\sum_{k=0}^{n-1}\frac{1}{nx+k}-\sum_{x\in S}\frac{1}{nx},\\ f(n)+\frac{1}{n}f(n) &= \sum_{k=1}^{n-1}\frac{1}{k}+\sum_{x\in S}\sum_{k=0}^{n-1}\frac{1}{nx+k}. \end{aligned}$$

Now for every positive integer b,

$$\frac{1}{b}=\log\left(\frac{b+1}{b}\right)+\frac{1}{2}\left(\frac{1}{b}-\frac{1}{b+1}\right)+O\left(\frac{1}{b^3}\right).$$

Putting $b=nx+k$ $(k=0,1,\cdots,n-1)$ and summing, we get

$$\sum_{k=0}^{n-1}\frac{1}{nx+k}=\log\left(\frac{x+1}{x}\right)+\frac{1}{2n}\left(\frac{1}{x}-\frac{1}{x+1}\right)+O\left(\frac{1}{n^2x^3}\right).$$

Hence

$$\begin{aligned}\frac{f(n)}{n} &= -\sum_{x\in S}\frac{1}{x}+\sum_{k=1}^{n-1}\frac{1}{k}+\sum_{x\in S}\log\left(\frac{x+1}{x}\right)+\frac{1}{2n}\sum_{x\in S}\left(\frac{1}{x}-\frac{1}{x+1}\right)\\ &\quad+O\left(\frac{1}{n^2}\right)\sum_{x\in S}\frac{1}{x^3}\end{aligned}$$

$$= -\sum_{\substack{x\in S\\ x>n}} \frac{1}{x} + \sum_{x=1}^{n-1} \log\left(\frac{x+1}{x}\right) + \sum_{\substack{x\in S\\ x>n}} \log\left(\frac{x+1}{x}\right)$$

$$+ \frac{1}{2n}\sum_{x=1}^{n-1}\left(\frac{1}{x} - \frac{1}{x+1}\right) + \frac{1}{2n}\sum_{\substack{x\in S\\ x>n}}\left(\frac{1}{x} - \frac{1}{x+1}\right) + O\left(\frac{1}{n^2}\right)$$

$$= \log n - \sum_{\substack{x\in S\\ x>n}} \left\{\frac{1}{x} - \log\left(\frac{x+1}{x}\right)\right\} + \frac{1}{2n} + O\left(\frac{1}{n^2}\right)$$

$$= \log n - \sum_{\substack{x\in S\\ x>n}} \left\{\frac{1}{2}\left(\frac{1}{x} - \frac{1}{x+1}\right) + O\left(\frac{1}{x^3}\right)\right\} + \frac{1}{2n} + O\left(\frac{1}{n^2}\right)$$

$$= \log n - \sum_{\substack{x\in S\\ n<x<n^2}} \frac{1}{2}\left(\frac{1}{x} - \frac{1}{x+1}\right) - \sum_{\substack{x\in S\\ x>n^2}} \frac{1}{2}\left(\frac{1}{x} - \frac{1}{x+1}\right)$$

$$+ \frac{1}{2n} + O\left(\frac{1}{n^2}\right)$$

$$= \log n - \sum_{x=n}^{n^2-1} \frac{1}{2}\left(\frac{1}{x} - \frac{1}{x+1}\right) + \sum_{x=1}^{n-1} \frac{1}{2}\left(\frac{1}{nx} - \frac{1}{nx+1}\right)$$

$$+ \frac{1}{2n} + O\left(\frac{1}{n^2}\right)$$

$$= \log n - \frac{1}{2n} + O\left(\frac{1}{n^2}\right) + O(1)\sum_{x=1}^{n-1} \frac{1}{n^2x^2} + \frac{1}{2n} + O\left(\frac{1}{n^2}\right)$$

$$= \log n + O\left(\frac{1}{n^2}\right).$$

Therefore $f(n) = n \log n + O(1/n)$, as required.

Also solved by D. M. Bloom, L. Carlitz, C. Gardner, and E. F. Schmeichel.

Three Hyperbolas and a Quadrangle

E 2206 [1969, 1138]. *Proposed by R. C. Lyness, Blackpool, England*

Each of three hyperbolas has for its foci a different two of three noncollinear points. Each pair of hyperbolas has a set of six common chords. Show that three pairs of chords, a pair from each set, form the six lines of a quadrangle. Show, further, that the minor axes of the hyperbolas bisect the sides of a triangle whose vertices are three of the four points of the quadrangle.

Solution by the proposer. Denoting distances from the foci by r_1, r_2, r_3, the

hyperbolic branches $r_1-r_2=c$, $r_3-r_2=a$ intersect where $r_1^2=r_2^2+2cr_2+c^2$ and $r_3^2=r_2^2-2ar_2+a^2$ and

$$l_2 \equiv -ar_1^2-(c-a)r_2^2+cr_3^2+ac(c-a)=0.$$

This last is the equation of a straight line, and the intersection of $r_1-r_2=-c$ and $r_3-r_2=-a$ lies on it. So it is a common chord.

Since $\sum b(c+a-b)l_2\equiv 0$, l_1, l_2, l_3 concur. If $-a$ is written for a in l_2, we have l_2', the chord joining the intersections of $r_1-r_2=\pm c$ and $r_3-r_2=\mp a$, and l_2', l_3' and l_1' concur. The four points of the quadrangle are (l_r, l_{r+1}'), $r=1, 2, 3$, and (l_1, l_2, l_3).

At the point of concurrency of l_1, l_2', l_3' we have $r_1^2-r_2^2=c(a+b)$, and at the point of concurrency of l_2, l_1', l_3' we have $r_1^2-r_2^2=-c(a+b)$. Hence the minor axis $r_1=r_2$ bisects the side $(l_1l_2', l_1'l_2)$.

Also solved by M. G. Greening (Australia) who establishes the result as a consequence of the following theorem which he proves using homogeneous coordinates: If three conics S_1, S_2, S_3 are inscribed respectively in the quadrilaterals A_2IA_3J, A_3IA_1J, A_1IA_2J, where no three of the points A_1, A_2, A_3, I, J are collinear, then there exist pairs of common chords of S_1, S_2; S_2, S_3; S_3, S_1 which form the six lines of a quadrangle.

A. W. Walker refers to E. H. Neville (*A focus-sharing set of three conics*, Math. Gazette, vol. 20, 1936, p. 182; also vol. 32, 1948, p. 184) who proves the first part of *E* 2206, and shows that if A, B, C are the diagonal points of the quadrangle, the directrices of one conic pass through B and C, those of another through C and A, and those of the third through A and B. The minor axes of the conics therefore bisect the sides of triangle ABC.

A Bound for Another Integral

E 2207 [1969, 1138]. *Proposed by Anon, Erewhon-upon-Wabash*

Suppose $f(x, y)$ vanishes on the boundary of the square S: $0\leqq x, y\leqq 1$, and that $|\partial^4 f/\partial x^2\partial y^2|\leqq B$. Prove that

$$\left|\iint_S f(x, y)dx\ dy\right| \leqq \frac{1}{144}B.$$

Solution by Ralph Jones, University of Wisconsin. Perhaps the simplest example of a function satisfying the conditions of the problem is $g(x, y)=x(1-x)y(1-y)$, for which

$$\frac{\partial^4 g}{\partial x^2\partial y^2}=4, \qquad \iint_S g=\frac{1}{36}=\frac{4}{144}\cdot$$

Since f vanishes on the boundary of S, $\partial^2 f/\partial y^2$ vanishes on the vertical sides of S. Four applications of integration by parts yield

$$\iint_S \frac{\partial^4 f}{\partial x^2\partial y^2}g=\iint_S f\frac{\partial^4 g}{\partial y^2\partial x^2}=4\iint_S f.$$

Thus

$$\left|\iint_S f\right| \leqq \frac{1}{4}\iint_S \left|\frac{\partial^4 f}{\partial x^2 \partial y^2} g\right| \leqq \frac{B}{4}\iint_S g = \frac{B}{144}.$$

Also solved by L. Carlitz, W. M. Causey, J. P. Fink, R. H. Garstang, Doug Hanto, M. Hioselihorn (Scotland), David Hoitsma, R. A. Horn, Simeon Reich (Israel), Steve Rodhe, E. F. Schmeichel, P. C. Sinha (U.S.S.R.), Sid Spital, Feets Uneven, E. J. Wilkins, Jr., P. H. Young, R. L. Young, and the proposer.

Several solvers note the similarity between this problem and E 2155 [1969, 1142], and are able to use the latter in their solution.

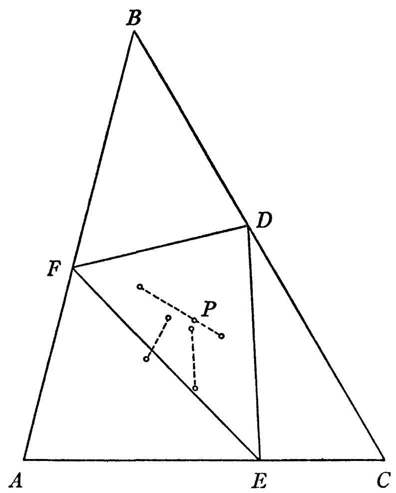
A Locus of Centroids

E 2209. *Proposed by M. S. Klamkin, Ford Scientific Laboratory*

Determine the locus of the centroids of all triangles similar to a given triangle and inscribed in another triangle.

Solution by Michael Goldberg, Washington, D.C. If a triangle of given shape grows so that the vertices trace fixed straight lines in the plane, then every point of the triangle will trace a straight line.

If the fixed straight lines are the sides of the triangle ABC, and the variable inscribed triangle is DEF, then its centroid P describes a straight line. However, there are some orientations for which the vertices of DEF cannot be confined to the straight line segments of the triangle ABC. Also, as DEF is turned, the motion of the vertices will change direction as they shift from one line of ABC to another line of ABC. Hence, the complete locus of the centroid consists of three straight line segments, shown in dotted lines in the figure.



If, in addition to direct symmetry, reflected symmetry is acceptable, then three more straight line segments are to be added to the locus of the centroid.

Also solved by Anders Bager (Denmark), Jordi Dou (Spain), Daniel Pedoe, A. W. Walker, Mark Yu, and the proposer.

Several solvers called attention to the well-known underlying theorem which can be found in Peterson's text, also Johnson's, and elsewhere.

Tea for Two

E 2210 [1970, 79]. *Proposed by R. N. Lloyd, University of Nottingham, England*

While the Dormouse lay sleeping in the middle of the table, the Mad Hatter and the March Hare sat down for tea. A number of places had been laid, and the Hatter and the Hare each moved alternately to a seat next to his old one. Each changed his direction of movement when, and only when, either the other had just changed his or else so as to avoid their both sitting on the same chair. Tea finished when both the Hatter and the Hare were again sitting in their initial positions. After how many moves was the tea-party over?

I. *Solution by R. W. Sielaff, Naperville, Ill. and G. P. Smith (undergraduate), MacMurray College, Jacksonville, Ill. (A composite by the editors.)* Let n be the number of places that were laid, $n>2$. If both the Mad Hatter and the March Hare move in the same direction, each will require n moves to regain his initial position since there will be no occasion for either to reverse direction. Consequently, the tea-party will be over after $2n$ moves.

Suppose then that they move in opposite directions and that there are k places between them in the direction of motion. To fix ideas, we assume without loss of generality that the Hatter moves first in the clockwise direction. There are three cases.

CASE I. Suppose k is even. After $k/2$ moves by each, they will be seated next to one another and each will reverse his direction of movement. After $k/2$ further moves by each, they will both have regained their initial places. Consequently the tea-party will be over after $2k$ moves.

CASE II. Suppose k is odd and n even. After $(k-1)/2$ moves by each, they will be separated by one place. After the next move by the Hatter, they will be seated next to one another, and each will reverse his direction of movement, the Hare first and then the Hatter. After $(n-2)/2$ additional moves by each, they will again be seated next to one another and each will reverse his direction of movement again, with the Hare reversing direction first and the Hatter following suit. After $(n-2)/2-(k-1)/2-1$ further moves by each, the Hatter will be in his initial position, having required a total of $(k-1)/2+1+(n-2)/2+(n-2)/2-(k-1)/2-1=n-2$ moves in all, whereas the Hare will require one further move to return to his initial place. That is, the Hare takes a total of $(k-1)/2+(n-2)/2+(n-2)/2-(k-1)/2=n-2$ moves in all. Consequently the tea-party will be over after $2n-4$ moves.

CASE III. Suppose both k and n odd. After $(k-1)/2$ moves by each, they

will be separated by one place. After the next move (by the Hatter), they will be seated next to one another and each will reverse his direction of movement, with the Hare reversing first and then the Hatter. After $(n-3)/2$ additional moves by each, they will again be separated by one place. After the next move (by the Hare), they will be seated next to one another, and it will be the Hatter's turn to move. After $(n-k-2)/2$ more moves by each, the Hatter and the Hare will each be one place in a clockwise direction from their original seats and, further, each will be moving in the direction he was moving orginally. This is because the Hatter will have made a total of $(k-1)/2+1+(n-k-2)/2=(n-1)/2$ moves in a clockwise direction and $(n-3)/2$ moves in a counterclockwise direction, whereas the Hare will have made a total of $(k-1)/2+(n-k-2)/2$ $=(n-3)/2$ moves in a counterclockwise direction, and a total of $(n-3)/2+1$ moves in a clockwise direction. Therefore n of these cycles will bring them both home moving in the original directions. However, during the last cycle, each one touches home at the same time while moving in the opposite direction, continues on for $(n-k-2)/2$ further moves, reverses, and returns in $(n-k-2)/2$ moves to finish the cycle. Hence they both return home for the first time (before the end of the last cycle) after a total of $(2n)((k-1)/2+1+(n-3)/2+(n-k-2)/2)$ $-4((n-k-2)/2)=2(n-1)(n-2)+2k$ moves.

II. *Solution by D. G. Beane and E. F. Schmeichel, the College of Wooster.* If the party starts with exactly two chairs separating the Hatter and the Hare in either direction, and the first move of each is onto these chairs, the party ends in four moves. Otherwise there are two possible outcomes. If the first move of each is clockwise or the first move of each is counterclockwise, the party ends in $2n$ moves (n denoting the number of places) since there is never a change of direction. If the first move of one is clockwise and the first move of the other is counterclockwise, they eventually occupy adjacent places. The next move of each is then back to the place which he previously occupied, and the following move of each again results in their occupying adjacent positions. This alternation will continue indefinitely, and the party may still be going on.

Also solved by Anders Bager (Denmark), Michael Goldberg, R. M. Ingle, S. Wu-Wei Liu, Norman Miller, Rik S. Moody, Paul Payne, Roger Weitzenkamp, and the proposer.

Editorial Comment. As the reader may have gathered from the two conflicting solutions, there was some confusion as to the proper interpretation of the "reversing rules" as given in the statement of the problem. The proposer (and your editors) were looking for a solution in the spirit of Solution I above; that is, the reversing rules should read (i) reverse if necessary to prevent a collision, and (ii) reverse in direct response to a reversal of type (i) by the other player. The statement of the problem unfortunately does not eliminate (iii) reverse in direct response to a reversal of type (ii) by the other player, (iv) reverse in direct response to a reversal of type (iii) by the other player, and so on. Allowing these, we are led to the alternating solutions as presented in Solution II.

The solvers were split about half and half between the two interpretations.

ADVANCED PROBLEMS

All solutions of Advanced Problems should be sent to J. Barlaz, Rutgers—The State University, New Brunswick, NJ 08903. Solutions of Advanced Problems in this issue should be typed (with double spacing) on separate, signed sheets and should be mailed before February 28, 1971. Contributors (in the United States) who desire acknowledgement of receipt of their solutions are asked to enclose self-addressed, stamped postcards.

An asterisk (*) means neither the proposer nor the editors supplied a solution.

5759. *Proposed by Robert Raphael, McGill University*

Let S be a commutative ring with unity, and let R be a subring of S containing the unity of S. Assume furthermore that S has no nilpotent elements (other than 0), and that each element of S satisfies a monic polynomial equation with coefficients from R. Recall that a ring is regular (in the sense of von Neumann) if for each x there is a y such that $x=x^2y$. Show that if R is regular then S is regular as well.

5760.* *Proposed by D. K. Kosaka, Dallas, Texas*

Without using the axiom of choice, or equivalent, decompose the unit interval $[0, 1]$ into an uncountable class of disjoint sets, each of which is uncountable and everywhere dense in the unit interval $[0, 1]$.

5761.* *Proposed by A. S. Adikesavan, Regional Engineering College, Tiruchirapalli, India*

Define $u(x)$ and $v(x)$ by the relation $\{u(x)\}^n+\{v(x)\}^n=1$. Where $n\geqq 2$ is an integer, $u(0)=0$, $v(0)=1$, and $du(x)/dx=v(x)$. Determine $u(x+y)$ in terms of $u(x)$, $u(y)$, $v(x)$ and $v(y)$.

5762. *Proposed by D. A. Zave, UNIVAC, Roseville, Minn.*

Let X be a real or complex normed linear space with norm $\|\cdot\|$. Let $a_i\neq 0$, $i=1, \cdots, n$ be scalars and let $p\geqq 1$. If $b_1, \cdots, b_n$ and s are fixed elements of X, compute

$$\inf\left\{\sum_{i=1}^{n}\|a_ix_i+b_i\|^p : x_1+\cdots+x_n=s\right\}.$$

5763.* *Proposed by J. T. Rosenbaum, University of Pittsburgh*

For any set S of reals, call the convergent series $\sum a_n$ an S series if for each $\epsilon>0$ there exists a sequence $\{I_n\}$ of open intervals covering S with $|I_n|\leqq\epsilon\, a_n$, $n=1, 2, \cdots$. Find a Cantor set series (see Problem 5665 [1970, 411]). Is there an S series for all S of measure 0? Is there a universal series?

5764. *Proposed by M. L. Glasser, Battelle Institute, Columbus, Ohio*

Evaluate the limit

$$L=\lim_{\alpha\to\pi}\sin\alpha\int_0^\infty\frac{\sinh(\gamma x)}{\sinh(\pi x)}\cdot\frac{dx}{\cosh x+\cos\alpha}, \qquad |\operatorname{Re}\gamma|<\pi+1.$$

SOLUTIONS OF ADVANCED PROBLEMS

Groups with a Cyclic Center

5689 [1969, 947]. *Proposed by Frank DeMeyer, Colorado State University*

Let G be a finite group with cyclic center Z and assume G/Z is abelian. Show $G/Z=H\times H$ (direct product) for some abelian group H.

Solution by C. R. B. Wright, University of Oregon. We prove the theorem: Let G be a finitely generated group with center Z and cyclic commutator group G'. If $G'\subseteq Z$, then $G/Z\simeq H\times H$ for some group H.

For each x in G the map $p_x\colon y\to[x, y]=x^{-1}y^{-1}xy$ is a homomorphism of G onto $[x, G]$ with kernel $C(x)$. Choose a and b in G to maximize the group $\langle[a, b]\rangle$. Then for some g, $[a, G]=Gp_a=\langle gp_a\rangle$, since G' is cyclic and therefore $[a, b]$ is in $\langle[a, g]\rangle$. Thus and similarly $\langle[a, b]\rangle=[a, G]=[b, G]$. Let $A=\langle a\rangle Z/Z$, $B=\langle b\rangle Z/Z$ and for each g in G let $g^\perp=C(g)/Z$.

Since p_a induces a homomorphism of G/Z onto $[a, G]$ with kernel $a^\perp$, and since $Gp_a=\langle b\rangle p_a$, $G/Z=a^\perp B$ and similarly $a^\perp=(a^\perp\cap b^\perp)A$. Thus $G/Z=(a^\perp\cap b^\perp)AB$.

If $|[a, b]|$ is infinite, then $A/A\cap b^\perp$ is infinite, and $A\cap b^\perp=1$. If $|[a, b]|$ is finite, then G' is finite and for each g, $A\cap g^\perp$ has index $|[a, g]|$ in A and hence contains $A\cap b^\perp$, the subgroup of index $|[a, b]|$ in A. In this case, too, $A\cap b^\perp=A\cap G^\perp=1$. By symmetry, $B\cap a^\perp=1$. In particular, $A\cap B=1$. Also by the modular law, $AB\cap a^\perp\cap b^\perp=A(B\cap a^\perp)\cap b^\perp=1$. Altogether, $G/Z=A\times B\times(a^\perp\cap b^\perp)$, with $A\simeq[a, G]=[b, G]\simeq B$.

Since the center of $C(a)\cap C(b)$ is central in G, $C(a)\cap C(b)$ satisfies the hypotheses of the theorem. Repetition of the above argument a finite number of times leads to the required decomposition of G/Z.

Also solved by Kevin Brown, M. G. Greening (Australia), C. Y. Tang, and the proposer.

Projective R-modules

5695 [1969, 1074]. *Proposed by Anon, Erewhon-upon-Wabash.*

Let $R=Z[a_1,a_2, a_3, b_1, b_2, b_3]$with the single relation $a_1b_1+a_2b_2+a_3b_3=1$. Let M be the R-module generated by x_1, x_2, x_3 with the single defining relation $b_1x_1+b_2x_2+b_3x_3=0$. Prove that M is not a free module, but is projective.

I. *Solution by G. J. Janusz, University of Illinois.* We interpret R as a noncommutative ring, and M as a left R-module. Then the problem is incorrect as stated. The result is true if the first of the defining relations is changed to read $b_1a_1+b_2a_2+b_3a_3=1$. I shall first show the result false in its original form and then present a proof of the corrected version.

Let F be the free left R-module with generators X_1, X_2, X_3 and let π be the homomorphism of F onto M such that $\pi(X_i)=x_i$. The definition of M implies Ker $\pi=RY$, where $Y=b_1X_1+b_2X_2+b_3X_3$. The module M is projective if and only if RY is an R-direct summand of F.

Suppose M is projective. Let α be some projection of F onto RY such that α is the identity on RY. Then $\alpha(X_i)=r_iY$ for some r_i in R. It follows that

$$Y=\alpha(Y)=\sum b_i\alpha(X_i)=\sum b_ir_iY.$$

After we express Y in terms of the X_i, we find that $(1-\sum b_jr_j)b_i=0$.

We now appeal to P. M. Cohn, *Some remarks on the invariant basis property*, Topology, v.5 (1966). The corollary on page 224 says this ring R has no zero divisors. (R is isomorphic to $U_{3,1}$ in Cohn's notation.) Thus $1=\sum b_j r_j$. This relation, however, cannot be a consequence of the single relation $\sum a_i b_i=1$; this argument can be made precise by using the notation of reduced forms below.

Now if the defining relation for R is changed to read $\sum b_i a_i=1$, define α by $\alpha(X_i)=a_iY$. Then $\alpha^2=\alpha$ and $\operatorname{Im}(\alpha)=RY$, hence RY is an R-direct summand of F, and M is projective over R.

To show that M is not free, we must again refer to Cohn's paper for the assertion that the ring R has the property that the number of free generators of a free R-module is an invariant of the module. Thus we suppose that M is free, and after replacing M with an isomorphic copy we find $D=M\oplus RY$ and M is free on two generators Y_1, Y_2. We know $M=\operatorname{Ker}\alpha$, where $\alpha(X_i)=a_iY$, so M is generated by the elements X_i-a_iY. If we write Y_3 for Y, then we obtain equations

$$X_i=\sum r_{ij}Y_j,\qquad Y_i=\sum s_{ij}X_j,$$

where all subscripts range from 1 to 3, r_{ij} and s_{ij} are in R, and $r_{i3}=a_i$, $s_{3i}=b_i$. Let A and B denote the matrices $\|r_{ij}\|$ and $\|s_{ij}\|$ respectively. Then $AB=BA=I$.

To proceed from here it will be necessary to express elements of R in some normal form. Any element of R is a sum of monomials involving the a_i and b_i, each multiplied by some integer coefficient. An element r will be called *reduced* if no monomial with nonzero coefficients contains the product b_1a_1. By using the results of J. Shepherdson: *Inverses and zero divisors in matrix rings*, Proc. London Math. Soc., 1 (1951) 71–85, it can be shown that each element is equal to a reduced element (eliminate b_1a_1 by using $\sum b_i a_i=1$) and this expression is unique.

As a consequence of $BA=I$, it follows that

$$0=\sum_i b_i r_{ik}\qquad\text{for } k=1\quad\text{or}\quad 2.$$

We assume all terms r_{ij} are in reduced form; then $b_2r_{2k}+b_3r_{3k}$ is also reduced, and hence b_1r_{1k} must not be reduced. For if b_1r_{1k} were also reduced, then so would be the sum; but a reduced term b_1r_{1k} cannot equal a reduced expression $-b_2r_{2k}-b_3r_{3k}$ unless all terms are zero, whereas the r_{ik} cannot all be zero.

Since b_1r_{1k} is not reduced, we find $r_{1k}=a_1w_{1k}$ for some w_{1k} in R. It now follows that the matrix A can be written as a product DC where $D=\operatorname{diag}\{a_1,1,1\}$ and C has the same second and third row as A, while row 1 of C is $(w_{11}, w_{12}, 1)$. Now the equation $DCB=1$ implies that there is some (reduced) element d in R such

that $a_1d=1$. This is impossible since whenever d is reduced, so is a_1d. This contradiction is a consequence of the assumption that M is free.

II. *Solution by the proposer*. We are dealing with modules over a commutative ring. Let F be the free R-module with basis X_1, X_2, X_3, and set

$$Y_0 = b_1X_1 + b_2X_2 + b_3X_3, \qquad Y_1 = a_2X_3 - a_3X_2,$$
$$Y_2 = a_3X_1 - a_1X_3, \qquad Y_3 = a_1X_2 - a_2X_1.$$

Let $G=RY_0\subseteq F$ and $H=RY_1+RY_2+RY_3\subseteq F$. Since $a_1Y_0=X_1+b_2Y_3-b_3Y_2$, etc., $G+H=F$. If $c_0Y_0=c_1Y_1+c_2Y_2+c_3Y_3\in G\cap H$, then $c_0b_1=c_2a_3-c_3a_2$, $c_0b_2=c_3a_1-c_1a_3$, $c_0b_3=c_1a_2-c_2a_1$. Multiply by a_1, a_2, a_3 and sum: $c_0=0$; hence $c_0Y_0=0$, $G\cap H=0$, $F=G\oplus H$. But $M\approx F/G\approx H$, hence M is a direct summand of a free module and therefore is a projective module. (See H. Flanders, Trans. A.M.S., 145 (1969) p. 366 for another proof.)

It seems much harder to prove algebraically that M is not free. In Flanders this is done by first identifying $a_i=b_i$ and the tensoring into the reals. The assertion then follows from the nonexistence of a nowhere zero (algebraic) vector field on S^2. A completely algebraic proof would be desirable.

Note that the module $M=Rx_1+Rx_2$ is free; here $b_1x_1+b_2x_2=0$, $R=Z[a_1, a_2, b_1, b_2]$, $a_1b_1+a_2b_2=1$. A basis is $y=a_2x_1-a_1x_2$. The analogous case $M=Rx_1+ \cdots +Rx_4$ is open. Note that the specialization

$$R = Z[a_0, a_1, a_2, a_3], \qquad a_0^2 + \cdots + a_3^2 = 1,$$
$$M = Rx_0 + \cdots + Rx_3, \qquad a_0x_0 + \cdots + a_3x_3 = 0$$

yields a free module M. The usual quaternion argument (S^3 is parallelizable) gives the basis

$$y_1 = - a_1x_0 + a_0x_1 - a_3x_2 + a_2x_3, \qquad y_2 = - a_2x_0 + a_3x_1 + a_0x_2 - a_1x_3.$$
$$y_3 = - a_3x_0 - a_2x_1 + a_1x_2 + a_0x_3.$$

Integral of a Uniform Sum

5700 [1969, 1074]. *Proposed by R. E. Chandler and R. A. Struble, North Carolina State College*

Let $\{x_n\}_{n=1}^\infty$ be an enumeration of the rational numbers in (0, 1). For $x\in(0, 1)$ define $f(x)=\sum 1/2^n$, where the summation is over all n for which $x_n<x$. Evaluate $\int_0^1 f(x)dx$.

Solution by W. H. Ruckle, Clemson University. For each n let

$$g_n(x) = \begin{cases} 0 & x \leqq x_n \\ 1/2^n & x > x_n. \end{cases}$$

Then $\sum_{n=1}^\infty g_n$ converges uniformly to f. So

$$\int_0^1 f(x) = \sum_{n=1}^\infty \int_0^1 g_n(x)dx = \sum_{n=1}^\infty (1 - x_n)/2^n.$$

Also solved by R. G. Bilyeu, David Burdick, David Burman, L. Carlitz, W. D. Causey, Red Cougar, D. Ž. Djoković, R. J. Driscoll, G. B. Feissner, Neal Felsinger, Robert Fefferman, N. J. Fine, Addison Fischer, T. E. Gantner, Leon Gerber, Rudolf Gorenflo (Germany), M. L. T. Hautus (Germany), D. A. Hejhal, G. A. Heuer, M. Hirshorn, K. E. Hirst (England), O. P. Kapoor & G. D. Lakhani (India), J. A. Kelingos, M. L. Klasi, R. A. Kopas, Frank Kost & William Weller, J. R. Kuttler, E. S. Langford, Douglas Lind, D. H. Lorenz & J. G. Simmonds, M. D. Mavinkurve (India), M. H. Moore, Roy Olson, Ed Packel, T. M. Phillips, J. T. Rosenbaum, C. L. Sabharwal & M. D. Fraser & J. T. McEwen, E. F. Schmeichel, A. C. Segal, D. Hammond Smith (England), Charles Vanden Eynden, M. S. Waterman, and the proposers.

Editorial Note. Hejhal proves the result in the form of the following abstraction: Let $\{a_n\}_{n=1}^{\infty}$ be a sequence in (0, 1) and let μ be a totally finite Borel measure on (0, 1). Set $Q(x)=\mu(x, 1)$, $0\leqq x\leqq 1$. Let $\sum_{n=1}^{\infty} g_n$ be a real, absolutely convergent infinite series. Finally, set $F(x)=\sum g_n$, where the summation is over all n such that $a_n<x$, and $0\leqq x\leqq 1$. Then:

(i) Q is monotonic decreasing, right continuous, and Borel measurable;

(ii) F is bounded and Borel measurable; and

(iii) $$\int_0^1 F d\mu = \sum_{n=1}^{\infty} g_n Q(a_n).$$

Smith, Driscoll, Lorenz, Simmonds, and Waterman note that the set of possible values of the integral is dense in (0, 1). But an unanswered question is whether the set of values consists of every number in (0, 1).

A Diophantine Equation

5703 [1969, 1152]. *Proposed by Erwin Just, Bronx Community College*

If $n>1$ and k is any integer, can there exist solutions to the Diophantine equation:

$$\sum_{i=1}^{2^{n+1}-1} (x_i^{2^n} - y_i^{2^n}) = (2k+1)2^{n+1}?$$

Solution by E. F. Schmeichel, Itasca, Illinois. We answer negatively. (All congruences are modulo 2^{n+2}.) It is easily verified that $a^{2^n}\equiv 0$ or 1 according as a is even or odd. So if all the x_i, y_i are integers, we have

$$x_i^{2^n} - y_i^{2^n} \equiv -1, 0, \text{ or } 1 \text{ for all } i.$$

Thus

$$S = \sum_{i=1}^{2^{n+1}-1} (x_i^{2^n} - y_i^{2^n}) \equiv d,$$

where $|d|\leqq 2^{n+1}-1$. But this means

$$k\cdot 2^{n+2} - (2^{n+1}-1) \leqq S \leqq k\cdot 2^{n+2} + (2^{n+1}-1)$$

for some integer k, or $(2k-1)2^{n+1}<S<(2k+1)2^{n+1}$. Since $2k-1$ and $2k+1$ are consecutive odd integers, no integer solution for the equation is possible.

Also solved by Merrill Barnebey, Simeon Reich (Israel), David Spear, Charles vanden Eynden, and the proposer.

A Linear Differential Equation

5705 [1969, 1152]. *Proposed by M. S. Klamkin, Ford Scientific Laboratory*
Find the general solution of the differential equation

$$[xD^{n+1} + 2nD^n - xD - n]y = 0.$$

Solution by Robert Heller, Mississippi State University. It can be shown by induction that $xD^n + nD^{n-1} = D^n x$, from which it follows that the given equation may be written $(D^n - 1)(xD + n)y = 0$. Hence

$$(xD + n)y = \sum_{k=1}^{n} c_k e^{a_k x},$$

where $a_1, \cdots, a_n$ are the n distinct nth roots of unity. Multiplication by x^{n-1} gives

$$D(x^n y) = \sum_{k=1}^{n} c_k x^{n-1} e^{a_k x}.$$

Repeated integration by parts shows that

$$x^n y = c_0 + \sum_{k=1}^{n} \left[c_k e^{a_k x} \sum_{p=1}^{n} (-1)^{p-1} \frac{(n-1)!}{(n-p)!} (a_k x)^{n-p} \right].$$

Taking limits of both sides as $x \to 0$, we see that although solutions y may exist on $(-\infty, \infty)$, such solutions are not expressed with precisely $n+1$ arbitrary constants. On $(-\infty, 0)$ or on $(0, \infty)$, we have the general solution

$$y = c_0 x^{-n} + x^{-n} \sum_{k=1}^{n} \left[c_k e^{a_k x} \sum_{p=1}^{n} (-1)^{p-1} \frac{(n-1)!}{(n-p)!} (a_k x)^{n-p} \right].$$

Also solved by I. N. Baker (England), M. T. Bird, Peter Brady, Jr., D. R. Breach (New Zealand), J. H. E. Cohn (England), Michael Deakin (New Guinea), Robert Desko, J. Gillis (Israel), K. K. Gorowara, Melvin Henriksen, J. M. Horner, O. P. Lossers (Netherlands), D. E. Myers, B. S. Popov (Yugoslavia), C. L. Sabharwal, St. Olaf College Students, L. E. Ward, Sr., J. H. Webb (South Africa), J. E. Wilkins, Jr., and the proposer.

A Closure Property and Hamel Bases

5706 [1969, 1152]. *Proposed by J. H. B. Kemperman, University of Rochester*

Let H be a Hamel basis of a field R over a subfield Q. Show that for each $a \in R$, $a \neq 1$, there exists an element $x \in H$ with $ax \notin H$.

Solution by Leroy F. Meyers, Ohio State University. For fixed $a \neq 1$, let $x \in H$ imply $ax \in H$. For each $b = \sum_{x \in H} b_x x \in R$ with $b_x \in Q$ and $\{x \mid b_x \neq 0\}$ finite, set $f(b) = \sum_{x \in H} b_x \in Q$. Then f is linear over R, and $f(ab) = f(b)$ (since $aH \subset H$). With $x \in H$ and $b = x/(a-1)$ we have $1 = f(x) = f(ab) - f(b) = 0$, a contradiction.

Also solved by A. K. Charnow, E. J. Eckert, K. E. Eldridge, N. J. Fine, D. A. Hejhal, M. L. Laplaza (Puerto Rico), M. A. Mavinkurve (India), P. van der Steen (Netherlands), Bertram Walsh, and the proposer.

Editorial Notes. The proposer and the solver note that this problem is an extension of 5538 (1968, 916). Eldridge and Walsh observe that this result does not depend on the commutativity of Q and hence remains true for division rings.

REVIEWS

EDITED BY KENNETH O. MAY

COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER

Beginning with the January 1971 issue this section will be edited by J. Arthur Seebach Jr. and Lynn A. Steen with the collaboration of the Mathematics Departments at St. Olaf College and Carleton College.

All material for review should be sent to: Book Review Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

All unsigned material is written by the editors. A boldface **C** *in the margin indicates that a review is based in part on classroom use.*

EDITORIAL: BOOK PRICES

A dramatic increase in book prices during the last two decades is evident to everyone. This has brought windfall profits on books already produced in an era of lower costs. But on the whole, book prices seem to have increased pretty much in keeping with the general inflation. A tabulation of Telegraphic Reviews for 1967–1969 shows that prices have not increased appreciably during this three year period. Indeed the average list price per page has remained 1.6¢ for paperbacks, 2.4¢ for elementary (first two years) books, and 3.6¢ for advanced books. This is about twice the levels of fifteen years ago.

The variations in prices among publishers are dramatic, and exasperating. For example, paperbacks sell from as little as 0.6¢ per page to as high as 13.2¢. Elementary texts (first two college years) vary from 1.2 to 7.5¢ and advanced books from 0.5 to 8.7¢. The variations are by no means always related to quality (often photo-offset of type script is priced the highest) or sales. The tremendous variation for paperbacks is due to the recent tendency to publish very advanced treatises in paperback. Since the cost of the hard cover is negligible, the lower price of paperbacks is due to the expected large market rather than to the nature of the binding. Accordingly if a book with a small market is published in paperback, it still has to be priced high (though probably not above 5¢ even for sales of only a few hundred) in order to avoid loss. The variations in prices among publishers are to some extent correlated with the type of book published, and therefore it is not fair to single out particular publishers for praise or blame on the basis of their average prices per page. However, as a general rule it appears that the well-established companies producing the books of highest quality are not those charging the highest prices. Indeed there appears to be no correlation whatever between price and quality of content or production.

THE AMERICAN MATHEMATICAL MONTHLY

(FOUNDED IN 1894 BY BENJAMIN F. FINKEL)

THE OFFICIAL JOURNAL OF

THE MATHEMATICAL ASSOCIATION OF AMERICA, INC.

VOLUME 77 NUMBER 10

CONTENTS

(continued on inside cover)

DECEMBER 1970

NOTICE TO AUTHORS

Articles should be typewritten on good quality paper, triple spaced with wide margins. Submit the original and a duplicate; keep a complete copy for protection against loss. Please follow the format used in current issues of the MONTHLY. The manuscript must make the author's intent clear to the printer. A paper accepted for publication by the editor but not acceptable to the printer will be returned to the author for retyping. Please use the American Mathematical Society *Manual for Authors of Mathematical Papers* as your guide and please conform to correct English usage. A *Manual for Monthly Authors* is available on request.

Prospective MONTHLY authors are advised to consult the Statement of Policy in the January 1969 issue, p. 2. Specialized research is usually unsuitable. Backlog: Main Articles 9 months, Math. Notes 9 months, Research Problems 7 months, Classroom Notes 7 months, Math. Education 8 months.

EDITORIAL CORRESPONDENCE AND MAIN ARTICLES: to HARLEY FLANDERS, American Mathematical Monthly, Tel Aviv University, Ramat Aviv, Israel (see Notice, p. 555); NOTES, etc.: to the corresponding Associate Editor; ADVERTISING CORRESPONDENCE: to RAOUL HAILPERN, Mathematical Association of America, SUNY at Buffalo, Buffalo, N. Y. 14214; CHANGE OF ADDRESS and SUBSCRIPTIONS: to A. B. WILLCOX, Mathematical Association of America, 1225 Connecticut Ave., N.W., Washington, D. C. 20036.

Annual dues for members of the Association (including a subscription to the American Mathematical Monthly) are $10.00. For nonmembers the subscription price is $15.00.

PUBLISHED BY THE ASSOCIATION at Washington, D. C., and Menasha, Wisconsin during the months of January, February, March, April, May, June–July, August–September, October, November, December.

Second-class postage paid at Washington, D. C., and additional mailing offices. Acceptance for mailing at special rate of postage provided for in the Act of February 28, 1925, embodied in Paragraph 4, Section 538, P. L. and R., authorized April 1, 1926.

PRINTED IN THE UNITED STATES OF AMERICA

VECTOR PROOFS IN SOLID GEOMETRY

M. S. KLAMKIN, Ford Motor Company

1. Introduction. In this paper, we give vector proofs of a number of theorems in solid geometry. Although, as to be expected, one can usually give synthetic proofs, one should always be prepared to use any alternate representations, especially if they are simpler. In many cases, it will turn out that not only is the vector approach simpler, but it is also more direct. Admittedly, however, in a great many other cases, the vector approach will not be quite as simple, but it still may be more direct. This does not mean that the author eschews the use of synthetic geometry, in fact the reverse is true. However, the main difficulty with elegant geometric solutions is that frequently there is no general method to indicate the first few proper steps or constructions which should be made. But once the first few proper steps have been made, the rest of the solution is usually very easy and one sees what is behind the problem. This is probably one of the reasons why analysts and algebraists tend to shun geometry; they have not given themselves enough practice to acquire the necessary geometric intuition to make the first few proper steps. Contrast this with analytic geometry. Here the first few steps are usually routine. You coordinatize everything in sight and then write down the appropriate equations. However, the subsequent equation solving can often be tedious, extremely difficult or unmanageable. On the other hand, vector proofs frequently appear to "lie between" synthetic and analytic proofs in regards to both simplicity and directness. The point to observe here is that one should be ready to use *any* representation, be it synthetic geometry, analytic geometry, analysis, vectors, algebra, etc., which lead to results. An exception to this, of course, is when one is learning some particular representation. This point is explored much further with many illustrative examples in [**1**] and [**2**].

Most of the theorems given here are known results. However, some of the proofs are believed to be new. Although some of them are not simpler than their synthetic counterparts, they still can be used to provide nontrivial and, we hope, interesting exercises to be used in classes on vector analysis. A number of these results are inequalities and, as to be expected, their proofs will involve either the triangle or Cauchy's inequality.

2. Angles in tetrahedra.

THEOREM 1: *The sum of the measures of any two face angles of a trihedral angle is greater than the measure of the third face angle* [3, p. 65].

Proof: Since $\cos x$ is decreasing in $[0, \pi]$, we equivalently have to show that

$$\cos AOC > \cos AOB \cos BOC - \sin AOB \sin BOC \text{ (see Fig. 1).}$$

Vectorially, this is equivalent to establishing the inequality

$$(\boldsymbol{B}\cdot\boldsymbol{B})(\boldsymbol{A}\cdot\boldsymbol{C}) > (\boldsymbol{B}\cdot\boldsymbol{C})(\boldsymbol{B}\cdot\boldsymbol{A}) - |\boldsymbol{B}\times\boldsymbol{C}|\,|\boldsymbol{B}\times\boldsymbol{A}|.$$

Since

$$(\boldsymbol{B}\cdot\boldsymbol{C})(\boldsymbol{B}\cdot\boldsymbol{A}) - (\boldsymbol{B}\cdot\boldsymbol{B})(\boldsymbol{A}\cdot\boldsymbol{C}) = \boldsymbol{B}\cdot\{\boldsymbol{C}(\boldsymbol{B}\cdot\boldsymbol{A}) - \boldsymbol{B}(\boldsymbol{A}\cdot\boldsymbol{C})\}$$
$$= \boldsymbol{B}\cdot[\boldsymbol{A}\times(\boldsymbol{C}\times\boldsymbol{B})] = (\boldsymbol{B}\times\boldsymbol{A})\cdot(\boldsymbol{C}\times\boldsymbol{B}),$$

the inequality immediately follows. There is equality only if O is in the plane of A, B, C.

THEOREM 2. *The sum of the areas of any three faces of a tetrahedron is greater than the area of the fourth face* [4, p. 106].

Proof: Referring to Fig. 1, we have to show vectorially that

$$|\boldsymbol{A}\times\boldsymbol{B}| + |\boldsymbol{B}\times\boldsymbol{C}| + |\boldsymbol{C}\times\boldsymbol{A}| > |(\boldsymbol{A}-\boldsymbol{B})\times(\boldsymbol{C}-\boldsymbol{B})|.$$

Since $|(\boldsymbol{A}-\boldsymbol{B})\times(\boldsymbol{C}-\boldsymbol{B})| = |\boldsymbol{A}\times\boldsymbol{B}+\boldsymbol{B}\times\boldsymbol{C}+\boldsymbol{C}\times\boldsymbol{A}|$, the desired result follows from the triangle inequality. Again there is equality only if O is in the plane of A, B, C.

THEOREM 3. *In any trihedral angle, the sines of the dihedral angles are proportional to the sines of the opposite face angles* [3, p. 67].

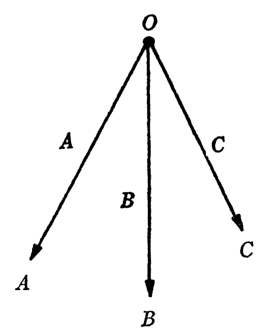


FIG. 1

Proof: If we let A, B, C, denote the dihedral angles containing the edges OA, OB, OC (which can now all be assumed to be of unit length), respectively, in Fig. 1, we have to show that

$$\frac{\sin A}{\sin BOC} = \frac{\sin B}{\sin COA} = \frac{\sin C}{\sin AOB}.$$

Since

$$\frac{\sin B}{\sin COA} = \frac{|(\boldsymbol{A}\times\boldsymbol{B})\times(\boldsymbol{B}\times\boldsymbol{C})|}{|\boldsymbol{A}\times\boldsymbol{B}|\cdot|\boldsymbol{B}\times\boldsymbol{C}|\cdot|\boldsymbol{C}\times\boldsymbol{A}|} = \frac{[\boldsymbol{A}\cdot\boldsymbol{B}\times\boldsymbol{C}]}{|\boldsymbol{A}\times\boldsymbol{B}|\cdot|\boldsymbol{B}\times\boldsymbol{C}|\cdot|\boldsymbol{C}\times\boldsymbol{A}|}$$

is symmetric in $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, the result is established.

The synthetic proof given in [3] does not include the proportionality constant which enables one to compute the dihedral angles from the face angles. To accomplish this, we need an expression for $[\boldsymbol{A}\cdot\boldsymbol{B}\times\boldsymbol{C}]$ (a neater way but not quite so direct is given in [5, p. 35]) which gives the volume of a parallelepiped having $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, as three coterminal edges in terms of the face angles. To this end, we let

$$\begin{aligned}
\boldsymbol{A} &= \boldsymbol{i},\\
\boldsymbol{B} &= \boldsymbol{i}\cos AOB + \boldsymbol{j}\sin AOB,\\
\boldsymbol{C} &= p\boldsymbol{i} + q\boldsymbol{j} + r\boldsymbol{k}, \quad \text{where} \quad p^2+q^2+r^2 = 1.
\end{aligned}$$

Since

$$\begin{aligned}
\boldsymbol{A}\cdot\boldsymbol{C} &= \cos AOC = p,\\
\boldsymbol{B}\cdot\boldsymbol{C} &= \cos BOC = p\cos AOB + q\sin AOB,
\end{aligned}$$

$$\begin{aligned}
(1)\quad V = [\boldsymbol{A}\cdot\boldsymbol{B}\times\boldsymbol{C}] &= r\sin AOB = (1-p^2-q^2)^{1/2}\sin AOB\\
&= \left\{(1-\cos^2 AOC)\sin^2 AOB - (\cos BOC - \cos AOC\cos AOB)^2\right\}^{1/2}\\
&= \left\{1-\cos^2 AOB - \cos^2 BOC - \cos^2 COA\right.\\
&\qquad \left. + 2\cos AOB\cos BOC\cos COA\right\}^{1/2}.
\end{aligned}$$

Thus $\sin A = V/(\sin AOB \sin COA)$, etc.

The symmetrical function $V/2$ of the face angles of the given trihedral angle is called the norm of the sides of the spherical triangle ABC and is denoted by n in spherical trigonometry [6, p. 498]. Also, V or equivalently $2n$ is sometimes called the sine of the solid angle subtended by the trihedral angle.

Reciprocally, the face angles as functions of the dihedral angles are obtained by letting

$$\sin BOC = \lambda \sin A, \text{ etc.}$$

and solving for λ as a function of A, B, C. Whence,

$$\frac{\sin AOB}{\sin A} = \frac{\sin BOC}{\sin B} = \frac{\sin COA}{\sin C} = \frac{2N}{\sin A\sin B\sin C},$$

where $2N = \left\{1-\cos^2 A - \cos^2 B - \cos^2 C - 2\cos A\cos B\cos C\right\}^{1/2}$. N is called the norm of the angles of the spherical triangle ABC. Alternate derivations by means of spherical trigonometry for the latter formulae are given in [6, pp. 497, 503].

Since V ranges between 0 and 1, we also have the following inequality:

$$3 \geqq \sin^2\theta_1 + \sin^2\theta_2 + \sin^2\theta_3 + 2\cos\theta_1\cos\theta_2\cos\theta_3 \geqq 2,$$

where

$$\pi \geqq \theta_i \geqq 0, \qquad \theta_1+\theta_2+\theta_3 \leqq 2\pi,$$

$$\theta_1+\theta_2 \geqq \theta_3, \qquad \theta_2+\theta_3 \geqq \theta_1, \qquad \theta_3+\theta_1 \geqq \theta_2.$$

The upper bound is achieved for $\theta_i = \pi/2$ and the lower bound when the sum of two angles equals the third angle.

A determinantal representation for V^2 is

$$V^2 = \begin{vmatrix} 1 & \cos AOB & \cos COA \\ \cos AOB & 1 & \cos BOC \\ \cos COA & \cos BOC & 1 \end{vmatrix}.$$

The latter form is equivalent to the following result given in [7, p. 48]:

THEOREM 4. *If α, β, γ are the angles between each pair of the triad of directions OA, OB, OC, then*

$$\begin{vmatrix} 1 & \cos\gamma & \cos\beta \\ \cos\gamma & 1 & \cos\alpha \\ \cos\beta & \cos\alpha & 1 \end{vmatrix} = \sin^2\gamma \sin^2\phi,$$

where ϕ is the angle OC makes with the plane OAB.

Proof: Referring to the previous figure again, $\sin\phi = (A\times B\cdot C)/|A\times B|$ or $V^2 = \sin^2\phi \sin^2 AOB$.

THEOREM 5. *If the edges of the base of a tetrahedron are a, b, c, and each of the lateral edges is equal to d, then the volume v of the tetrahedron is given by* [4, p. 108]

$$v = \{16p(p-a)(p-b)(p-c)d^2 - a^2b^2c^2\}^{1/2},$$

where $2p = a+b+c$.

Proof: It follows from (1), that $v = Vd^3/6$. By substituting $\cos BOC = 1 - a^2/2d^2$, etc., in V in (1) and simplifying we get the desired result.

THEOREM 6: *If a line makes congruent angles with each of three lines in a plane, the line is perpendicular to the plane* [3, p. 45].

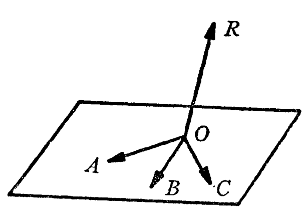


FIG. 2

Proof: If A, B, C denote unit vectors along the three lines in the plane and R is a vector along the other line, we have to show equivalently that (see Fig. 2)

$$R\cdot A = R\cdot B = R\cdot C \Rightarrow R\cdot A = R\cdot B = 0.$$

Let $R\cdot A = \lambda$. Then since A, B, C are coplanar, $C = aA + bB$ and

$$R\cdot C = (a+b)\lambda = \lambda \text{ or } (a+b-1)\lambda = 0.$$

But $a+b=1$ is impossible, since this implies that A, B, C are collinear. Thus $\lambda=0$. (A simple geometric proof follows by considering the perpendicular bisecting planes of AB and BC. Their intersection which is normal to the plane contains both O and R.)

3. Altitudes. The next set of theorems (7–12) relate to the altitudes of a tetrahedron. While the proofs of all of them are not as simple as the corresponding synthetic ones, they may be more appealing (especially to non-geometers) by virtue of their directness. All of them except Theorem 12 are given in [**4**, pp. 61–65].

All the proofs will refer to Fig. 3. Here $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$, etc., will denote vectors from a common origin to the points A, B, C, D, etc.

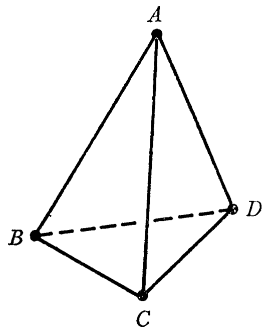


FIG. 3

THEOREM 7. *If a pair of opposite edges of a tetrahedron are rectangular (normal to each other), the two altitudes of the tetrahedron issued from the end of each of these two edges intersect.*

Proof: Equivalently, we have to show that given $(\mathbf{A}-\mathbf{B})\cdot(\mathbf{C}-\mathbf{D})=0$, there exists a vector $\mathbf{H}_1$ such that

$$(\mathbf{H}_1 - \mathbf{A})\cdot(\mathbf{B} - \mathbf{C}) = (\mathbf{H}_1 - \mathbf{A})\cdot(\mathbf{C} - \mathbf{D}) = 0, \tag{2}$$

$$(\mathbf{H}_1 - \mathbf{B})\cdot(\mathbf{C} - \mathbf{D}) = (\mathbf{H}_1 - \mathbf{B})\cdot(\mathbf{D} - \mathbf{A}) = 0. \tag{3}$$

If H is any point on h_a, then (2) is satisfied by $\mathbf{H}_1=\mathbf{A}+\lambda(\mathbf{H}-\mathbf{A})$ for any λ. Since

$$(\mathbf{A} - \mathbf{B})\cdot(\mathbf{C} - \mathbf{D}) = 0 \Rightarrow (\mathbf{H}_1 - \mathbf{B})\cdot(\mathbf{C} - \mathbf{D}) - (\mathbf{H}_1 - \mathbf{A})\cdot(\mathbf{C} - \mathbf{D}) = 0,$$

the first part of (3) is satisfied. To satisfy the second part of (3), we must have

$$(\mathbf{A} - \mathbf{B})\cdot(\mathbf{D} - \mathbf{A}) + \lambda(\mathbf{H} - \mathbf{A})\cdot(\mathbf{D} - \mathbf{A}) = 0.$$

Since $(\mathbf{H}-\mathbf{A})\cdot(\mathbf{D}-\mathbf{A})\neq 0$ a unique λ exists.

THEOREM 8 (converse of Th. 7). *If two altitudes of a tetrahedron intersect, the edge joining the two vertices from which these altitudes issue is perpendicular to the opposite edge of the tetrahedron.*

Proof: We wish to show equivalently that

$$\left.\begin{array}{ll}(4) & (H-A)\cdot(C-D)=(H-A)\cdot(B-C)=0\\(5) & (H-B)\cdot(C-D)=(H-B)\cdot(A-C)=0\end{array}\right\}\Rightarrow(A-B)\cdot(C-D)=0.$$

On subtracting the first part of (4) from the first part of (5), we get the desired result. Since we did not use the second parts of (4) and (5), we can make the stronger statement:

THEOREM 8'. *If two lines from A and B are normal to CD and intersect, then $AB\perp CD$.*

COROLLARY. *If two altitudes of a tetrahedron intersect, the remaining two altitudes intersect.*

THEOREM 9. *If two pairs of opposite edges of a tetrahedron are rectangular, then the remaining pair of opposite edges is also rectangular.*

Proof: We wish to show equivalently that

$$\left.\begin{array}{ll}(6) & (A-B)\cdot(C-D)=0\\(7) & (B-C)\cdot(D-A)=0\end{array}\right\}\Rightarrow(A-B)\cdot(B-D)=0.$$

The result immediately follows by subtracting (6) from (7).

THEOREM 10. *If three altitudes of a tetrahedron are concurrent, then the four altitudes are concurrent.*

Proof: Our first proof is similar to the vector proof for the concurrency of three altitudes of a triangle given in [8, p. 32]. A normal to the plane of A, B, C, is given by

$$\lfloor A\times B+B\times C+C\times A\rfloor.$$

Also,

$$A\times(B\times C)+B\times(C\times A)+C\times(A\times B)=0.$$

Using the latter identity, it follows immediately that

$$\begin{array}{ll} & (H-A)\times(B\times C+C\times D+D\times B)\\ (8) & +(H-B)\times(C\times A+A\times D+D\times C)\\ & +(H-C)\times(D\times A+A\times B+B\times D)\\ & +(H-D)\times(A\times C+C\times B+B\times A)=0\end{array}$$

is also an identity. By choosing H to be the point of concurrency of three of the altitudes, our result follows.

Although this proof is rather neat, it admittedly is not very direct. Consequently, we now give a more direct one.

We wish to show equivalently that

$$
(9)\quad \left.\begin{aligned}
(H-A)\cdot(B-C) &= (H-A)\cdot(C-D) = 0\\
(H-B)\cdot(C-D) &= (H-B)\cdot(D-A) = 0\\
(H-C)\cdot(D-A) &= (H-C)\cdot(A-B) = 0
\end{aligned}\right\}
$$
$$\Rightarrow (H-D)\cdot(A-B) = (H-D)\cdot(B-C) = 0.$$

As in Theorem 8, we then have $(A-B)\cdot(C-D)=(B-C)\cdot(D-A)=0$. Since

$$C - D = (H - D) - (H - C),$$
$$(H - D)\cdot(A - B) = (H - C)\cdot(A - B) = 0.$$

Similarly, $(H-D)\cdot(B-C)=0$.

THEOREM 11. *If $ABCD$ is a tetrahedron whose altitudes are concurrent in the point H (orthocenter), then each of the five points is the orthocenter of the tetrahedron formed by the other four points.*

Proof: Follows immediately from (9).

THEOREM 12: *If one altitude of a tetrahedron intersects two other altitudes, then the four altitudes are concurrent.*

Proof: It follows from Theorems 8 and 9 that

$$(B - C)\cdot(A - D) = (A - C)\cdot(B - D) = (A - B)\cdot(C - D) = 0.$$

If H denotes the point of intersection of h_a and h_b, then

$$(H - A)\cdot(B - C) = (H - A)\cdot(B - D) = 0,$$
$$(H - B)\cdot(A - C) = (H - B)\cdot(A - D) = 0.$$
$$(B - D)\cdot(A - C) = (B - D)\cdot(H - C) - (B - D)\cdot(H - A) = 0,$$
$$(A - D)\cdot(B - C) = (A - D)\cdot(H - C) - (A - D)\cdot(H - B) = 0.$$

Whence, $(H-C)\cdot(A-D)=(H-C)\cdot(B-D)=0$, implying that h_a, h_b and h_c are concurrent. Then by Theorem 10, all the altitudes are concurrent.

4. Tangencies. The next two theorems, which on first glance may seem unrelated, will be shown to be essentially equivalent.

THEOREM 13. *If an ellipsoid is tangent to the four edges of a skew quadrilateral, then the points of tangency are coplanar* [5, p. 180].

THEOREM 14. *If four spheres touch in succession, each one touching two others (the number of external contacts being even), then the four points of tangency lie on a circle* [5, p. 44].

We first establish Carnot's theorem and its converse [4, p. 111].

THEOREM 15. *If P, Q, R, S are four coplanar points on the four sides, respectively, of a skew quadrilateral $ABCD$, then*

$$\frac{AP}{PB}\cdot\frac{BQ}{QC}\cdot\frac{CR}{RD}\cdot\frac{DS}{SA} = 1 \tag{10}$$

and conversely (here AP is taken to be positive if P lies on the ray AB, otherwise it is negative). (See Fig. 4.)

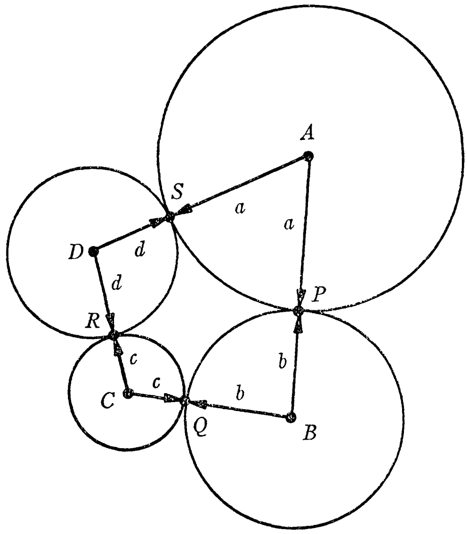


FIG. 4

Proof: Since P is on line AB etc., we have

$$\boldsymbol{P} = a\boldsymbol{A} + (1-a)\boldsymbol{B}, \qquad \boldsymbol{Q} = b\boldsymbol{B} + (1-b)\boldsymbol{C},$$
$$\boldsymbol{R} = c\boldsymbol{C} + (1-c)\boldsymbol{D}, \qquad \boldsymbol{S} = d\boldsymbol{D} + (1-d)\boldsymbol{A}.$$

Since P, Q, R, S are coplanar, there must exist four constants p, q, r, s, where $p+q+r+s=0$, $pqrs\neq 0$, such that

$$p\boldsymbol{P} + q\boldsymbol{Q} + r\boldsymbol{R} + s\boldsymbol{S} = \boldsymbol{0}.$$

Since the origin of the vectors is taken outside the 3-space of the quadrilateral, $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}$ are linearly independent. Thus the coefficients of $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}$ in the latter equation are all zero, i.e.,

$$p(1-d) + qa = 0, \qquad q(1-a) + rb = 0,$$
$$r(1-b) + sc = 0, \qquad s(1-c) + pd = 0.$$

In order that this latter set of equations be consistent, it is necessary that

$$(1-a)(1-b)(1-c)(1-d) = abcd.$$

On solving for p, q, r in terms of s and substituting back in $p+q+r+s$, it follows

after some simplification that the condition is also sufficient. Since

$$AP = (1 - a)\,|\boldsymbol{A} - \boldsymbol{B}|, \qquad PB = a\,|\boldsymbol{A} - \boldsymbol{B}|,$$

etc., we get the desired result.

For the converse theorem, we are given that (10) holds. Let the plane determined by P, Q, R, intersect AD in S'. Then by the previous theorem,

$$\frac{AP}{PB}\cdot\frac{BQ}{QC}\cdot\frac{CR}{RD}\cdot\frac{DS'}{S'A} = 1.$$

Whence S' coincides with S, and P, Q, R, S are coplanar.

We digress to sketch out an analytic geometry proof and a proof using centroids. Then we shall compare the proofs, including the synthetic one (using Menelaus' theorem) which was referred to in [4], to illustrate the differences in simplicity and directness as mentioned in the introduction.

For the analytic proof, we start out by coordinatizing the points A, B, C, D, i.e., let them be (x_i, y_i, z_i), where $i = 1, 2, 3, 4$. Points P, Q, R are then chosen arbitrarily on the lines AB, BC, and CD, respectively. S is then determined as the intersection of the plane through P, Q, R with the line DA. Finally, the desired relation is verified.

For the centroid proof, assume that masses

$$M_a = \frac{k}{AP}, \qquad M_b = \frac{k}{BP}, \qquad M_c = \frac{l}{CR}, \qquad M_d = \frac{l}{DR}$$

are placed at the points A, B, C, D, respectively, such that k, l satisfy

$$k\frac{AS}{AP} = l\frac{DS}{DR}\,.$$

These masses have been so chosen that the centroid of M_a and M_b is at P, the centroid of M_c and M_d is at R, and the centroid of M_a and M_d is at S. It now follows immediately that the centroid G of the four masses is on the segment $\overline{PR}$. If the centroid of M_b and M_c is at Q' on $\overline{BC}$, then G is also on SQ'. Thus, $\overline{BC}$ must intersect SQ' at G. Unless Q' and Q coincide A, B, C, D will be coplanar, contradicting the hypothesis. Then,

$$k\frac{BQ}{BP} = l\frac{CQ}{CR}$$

which together with the previous condition in k, l gives the desired result.

On comparing the different proofs it is seen that the analytic geometry one is very direct but involves a lot of arithmetic. The vector proof is somewhat less direct but the arithmetic is much less onerous. The computations in the centroid proof are few and easy but the proof requires the very indirect step of introducing appropriate masses and using centroids. The synthetic proof (see reference)

is also very easy provided that you invoke Menelaus' theorem. To any decent geometer, this would be a fairly direct step. However, to many others, it will not be.

We now return to the proofs of Theorems 13 and 14. Since tangency and planarity are preserved under affine transformations it suffices in Theorem 13 to prove the result for a sphere. Also, since tangents from a point to a sphere are congruent, the configuration we obtain is the same as in the first figure for the four possible configurations, Figs. 4–7 of Theorem 14.

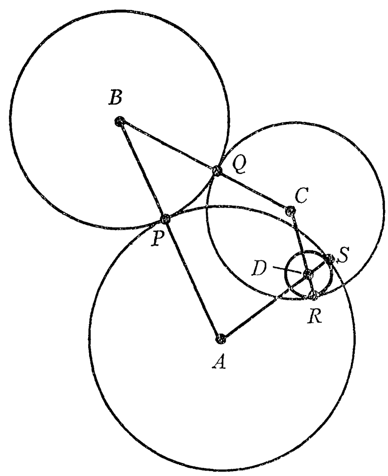


FIG. 5

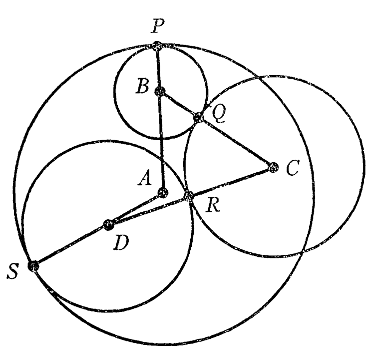


FIG. 6

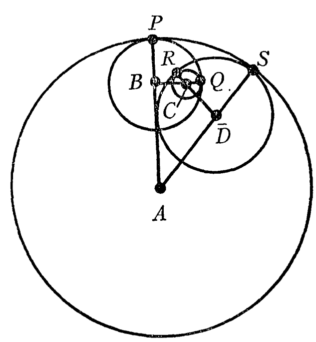


FIG. 7

In three of the cases (Figs. 5, 6, 7), some of the points of tangency P, Q, R, S do not lie in the interior of their corresponding segments AB, BC, CD, DA. This occurs when a pair of corresponding spheres are tangent internally, e.g., P in

Figs. 6, 7. However, in all the cases,

$$\frac{AP}{PB} = \pm\frac{a}{b}, \qquad \frac{BQ}{QC} = \pm\frac{b}{c}, \qquad \frac{CR}{RD} = \pm\frac{c}{d}, \qquad \frac{DS}{SA} = \pm\frac{d}{a},$$

where the number of minus signs is even. Then by the converse of Carnot's theorem, P, Q, R, S are coplanar. Since in Theorem 13, P, Q, R, S are on a sphere, they also must lie on a circle. To prove the circular part for Theorem 14 requires an extension of the proof that a sufficient condition (which is also necessary) for a plane quadrilateral $ABCD$ to have an inscribed circle is that $AB+CD=BC+DA$ [9, p. 28]. Our proof is geometric as we do not have a simple enough vectorial one. It will be in two parts, one for Figs. 4, 5 and the other for Figs. 6, 7.

Since P, Q, R, S have been shown to be coplanar, it suffices to show also that they are cospherical.

In Figs. 4, 5, $AB+CD=BC+DA$. Assuming without loss of generality that $AB \geqq DA$, then $BC \geqq CD$. Points X, Y are chosen on AB and BC, respectively, such that $AX=DA$, $YC=CD$ and then $XB=BY$. It is to be noted that although the configuration in Fig. 8 is taken from Fig. 4, it also applies to Fig. 5. The only change is that R and S are then exterior points of their corresponding segments. The perpendicular bisecting planes of DX, XY, YD will intersect in a line through the circumcenter of $\triangle DXY$ and perpendicular to its plane. By the symmetry of the configuration, each point on this perpendicular line will be equidistant from the four sides of $ABCD$ and also from P, Q, R, S. Thus, P, Q, R, S are cospherical and, additionally, the skew quadrilateral has a family of "inspheres."

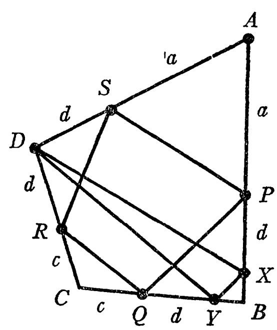


FIG. 8

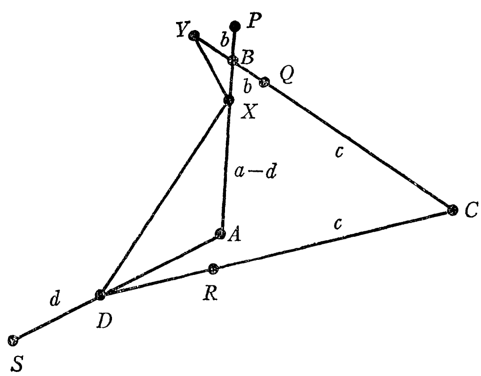


FIG. 9

In Figs. 6, 7, $AB+BC=CD+DA$. In the case of a plane quadrilateral, this will turn out to be the necessary and sufficient condition for the existence of a circle tangent to the four sides (2 internal and 2 external points of contact). Since the previous argument now applies also for Fig. 9, our proof is completed.

5. Higher dimensions. In the next four theorems, we consider n-dimensional figures. For problems in higher dimensions, the vector approach is particularly useful and often a necessity.

THEOREM 16. *Corresponding to any n-dimensional simplex, there exists a skew $(n+1)$-gon whose sides are congruent and parallel to the $n+1$ medians of the simplex. Furthermore, the volume of the simplex spanned by the $n+1$ vertices of the skew-gon is $(1+1/n)^{n-1}/n$ times the volume of the initial simplex. This generalizes the known results for a triangle and tetrahedron* [4, p. 54].

Proof: Let the $n+1$ vertices of the given simplex be $O, A_1, A_2, \cdots, A_n$ and let A_i denote the set of n linearly independent vectors from O to A_i. The volume of the given simplex is then

$$V = [A_1, A_2, \cdots, A_n]/n! = \frac{1}{n!}\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}.$$

Here $(a_{i1}, a_{i2}, \cdots, a_{in})$ denotes the rectangular components of A_i. The $n+1$ medians of the simplex are given by

$$nM = A_1 + A_2 + \cdots + A_n, \qquad M - \frac{n+1}{n} A_i \; (i = 1, 2, \cdots, n).$$

The sum of these $n+1$ vectors is zero and consequently they form a closed skew $(n+1)$-gon.

If we now choose the starting point of M as our origin, the n vectors to the n remaining vertices of the $(n+1)$-gon are

$$M,\ 2M - \lambda A_1,\ 3M - \lambda(A_1 + A_2), \cdots, nM - \lambda(A_1 + A_2 + \cdots + A_{n-1}),$$

where $\lambda = (n+1)/n$. The volume V' of this simplex is given by

$$V'/n! = [M, 2M - \lambda A_1, \cdots, nM - \lambda(A_1 + A_2 + \cdots + A_{n-1})].$$

By elementary operations on the determinant,

$$V'/n! = [M, \lambda A_1, \cdots, \lambda A_{n-1}] = [A_n/n, \lambda A_1, \lambda A_2, \cdots, \lambda A_{n-1}]$$

or

$$V'/n! = \frac{\lambda^{n-1}}{n}[A_1, A_2, \cdots, A_n].$$

It is known [10, p. 139] that in order that two points in the interior or on the boundary of a polygon be farthest apart, they must be two of the vertices that are farthest apart. The polygon need not be convex. An extension of this to n-dimensional polytopes is easily proved by means of the triangle inequality.

THEOREM 17. *If two points in the interior or on the boundary of a polytope are furthest apart, they must be two of the vertices which are furthest apart.*

Proof: We need only consider convex polytopes. For if the result is valid for the convex hull of the polytope, it is also valid for the polytope.

Now let $V_1, V_2, \cdots, V_n$ denote vectors from a common origin to the vertices of the polytope. Then $R=\sum \lambda_i V_i$ and $R'=\sum \lambda_i' V_i$, where $\lambda_i, \lambda_i' \geqq 0$, $\sum \lambda_i = \sum \lambda_i' = 1$ will denote two vectors from the common origin to two points within or on the boundary of the polytope. Using the triangle inequality repeatedly and the properties of λ_i and λ_i', we get the following sequence of inequalities:

$$\begin{aligned} |R'-R| &= \Big|\sum_i \lambda_i'(V_i-R)\Big| \leqq \sum_i \lambda_i' \,|V_i-R| \\ &\leqq \operatorname*{Max}_i |V_i-R| = \operatorname*{Max}_i \Big|\sum_j \lambda_j(V_j-V_i)\Big| \\ &\leqq \operatorname*{Max}_i \sum_j \lambda_j |V_j-V_i| \leqq \operatorname*{Max}_{i,j} |V_j-V_i|. \end{aligned}$$

For our next theorem, we consider a special case of "the optimal location of a warehouse" problem [11, p. 394]. Here we have to determine a point R which minimizes $\sum w_i|R-V_i|$, where V_i, are given points and w_i are given positive weights. If we set the appropriate derivatives equal to zero, the resulting equations are given by

$$\sum w_i \frac{R-V_i}{|R-V_i|} = 0$$

(a set of forces in equilibrium). It has been shown that the point R is unique. A minor difficulty arises if R coincides with one of the V_i's.

If the weights w_i are equal and the given points V_i correspond to the vertices of a regular polytope (in any dimension), then one expects intuitively from the symmetry of the configuration that R will correspond to the center of the polytope. We now give an extension of this which is established elementarily using the triangle inequality.

THEOREM 18. *Let V_i denote vectors from a common origin O to n given points on a unit sphere (of any dimension) with center O. If $\sum w_i V_i = 0$, where w_i are arbitrarily given positive weights, then*

$$\operatorname*{Min}_R \sum_i w_i |V_i-R| = \sum_i w_i.$$

Proof: By symmetry (see Fig. 10), $|V_i-R|=|rV_i-R/r|$. Whence,

$$\sum w_i |V_i-R| = \sum_i w_i |rV_i - R/r| \geqq \Big|\sum_i (rw_iV_i - w_iR/r)\Big| = \sum_i w_i$$

with equality, if and only if, $r=0$.

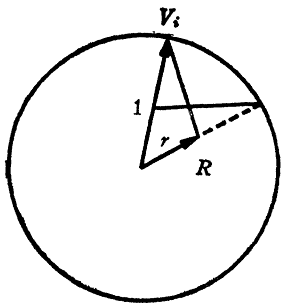


FIG. 10

Our last theorem is an elementary generalization of the convexity of ellipsoids of revolution. The proof via vectors is almost automatic. However, a simpler proof for general ellipsoids follows by transforming the ellipsoid into a sphere by an affine transformation (which preserves convexity).

THEOREM 19. *If $A_1, A_2, \cdots, A_r$ denote arbitrary points, not necessarily distinct, in E_n, then the region spanned by the points P satisfying the inequality $PA_1+PA_2+\cdots+PA_r \leqq k$ (constant) is convex.*

Proof: We first show that for any three collinear points P_1, Q, P_2, where Q is strictly between P_1 and P_2, that

$$\sum_{i=1}^{r} |\mathbf{Q}-\mathbf{A}_i| \leqq \max\left\{\sum_{i=1}^{r} |\mathbf{P}_1-\mathbf{A}_i|, \quad \sum_{i=1}^{r} |\mathbf{P}_2-\mathbf{A}_i|\right\}. \tag{11}$$

Here $\mathbf{A}_i$ denotes the vector from an origin O to the point A_i, etc.

Since Q is between P_1 and P_2, $\mathbf{Q}=\lambda\mathbf{P}_1+(1-\lambda)\mathbf{P}_2$ with $0<\lambda<1$. Then by the triangle inequality,

$$\begin{aligned} |\mathbf{Q}-\mathbf{A}_i| &= |\lambda(\mathbf{P}_1-\mathbf{A}_i)+(1-\lambda)(\mathbf{P}_2-\mathbf{A}_i)| \\ &\leqq \lambda|\mathbf{P}_1-\mathbf{A}_i| + (1-\lambda)|\mathbf{P}_2-\mathbf{A}_i|. \end{aligned}$$

Whence,

$$\sum_i |\mathbf{Q}-\mathbf{A}_i| \leqq \lambda\sum_i |\mathbf{P}_1-\mathbf{A}_i| + (1-\lambda)\sum_i |\mathbf{P}_2-\mathbf{A}_i|$$

with equality if and only if the points are collinear and also (11).

Now if P_1 and P_2 satisfy $\sum_i PA_i \leqq k$, so also does Q and thus the region is convex. For a generalization of Theorem 19, see [**12**].

References

1. M. S. Klamkin and D. J. Newman, The philosophy and applications of transform theory, SIAM Review, January 1961, pp. 10–36.

2. M. S. Klamkin, The teaching of mathematics so as to be useful, Educational Studies in Math., 1 (1968) 126–160.

3. L. Lines, Solid Geometry, Dover, New York, 1965.
4. N. Altschiller-Court, Modern Pure Solid Geometry, Macmillan, New York, 1935.
5. D. Y. Sommerville, Analytical Geometry of Three Dimensions, Cambridge University Press, Cambridge, 1947.
6. T. M. MacRobert and W. Arthur, Trigonometry, Part IV, Methuen, London, 1938.
7. A. J. McConnell, Applications of the Absolute Differential Calculus, Blackie and Son, Glasgow, 1946.
8. L. Brand, Vector and Tensor Analysis, Wiley, New York, 1948.
9. C. V. Durell and A. Robson, Advanced Trigonometry, Bell, London, 1953.
10. H. Rademacher and O. Toeplitz, The Enjoyment of Mathematics, Princeton University Press, N. J., 1957.
11. Problem 62–11, SIAM Review, Oct. 1962.
12. P. Hartman and F. A. Valentine, On general ellipses, Duke Math. J., 26 (1959) 373–385.

A PROOF OF THE NEWTON-COTES QUADRATURE FORMULAS WITH ERROR TERM

D. R. HAYES, University of Massachusetts, AND L. RUBIN, University of Hartford

1. Introduction. The Newton-Cotes quadrature formulas with error term are well known. A survey of recent textbooks on numerical analysis, however, seems to indicate that rigorous proofs of the standard expressions for the error terms are not nearly so well known as the expressions themselves. A relatively complete and rigorous proof can be found, for example, in Steffensen [3]; but the reader must first master a rather considerable amount of machinery. Further, Steffensen's method does not apply so well in the case of an even number of interpolation points as it does in that of an odd number, and additional difficulties must be surmounted. Our aim in this paper is to present a direct and elementary proof of these formulas with error term. It is hoped that the proof is simple enough that it can be presented to an undergraduate class in numerical analysis in three or four lectures. Besides a standard mean value theorem for integrals, we require no more mathematical background than is usually taught in undergraduate calculus courses. Our method applies equally well whether the number of interpolation points is odd or even. Further, although our main aim is a simplified proof, we also arrive at a slightly more general theorem than is sometimes stated, since we do not require any continuity assumptions for the derivatives of the function being integrated.

We now introduce some conventions and notation which will be used

---

David R.Hayes received his Duke Ph.D. in 1963 under L. Carlitz. He was assistant and associate professor at the Univ. of Tennessee before his present associate professorship at the Univ. of Mass. and spent an NSF postdoc. at Harvard. His research is mainly in analytic number theory.

Laurence Rubin received his Univ. of Tennessee M.A. under Hayes in 1965. Since, he has been at the Univ. of Hartford, Conn., and is presently at Nashville State Tech. He has held several NSF in-service awards. *Editor.*

throughout the paper. We assume given a positive integer n, and we define k by

$$k = \begin{cases} n/2 \text{ if } n \text{ is even} \\ (n-1)/2 \text{ if } n \text{ is odd.} \end{cases} \tag{1.1}$$

We also assume given a real valued function f which is defined and differentiable $2k+2$ times in some open interval containing the closed interval $[a, b]$. Further, we assume that the values of f are known explicitly at the $n+1$ points $a=x_0 < x_1 < \cdots < x_n = b$ which one obtains as endpoints when $[a, b]$ is partitioned into n **equal** subintervals. Let p be the **unique** polynomial of degree less than or equal to n which takes the same values as f at each of the points $x_0, \cdots, x_n$. We call p the **interpolating polynomial.** Set

$$A(x) = \prod_{i=0}^{2k} (x - x_i), \tag{1.2}$$

so that A is a polynomial of degree $n+1$ or n depending on whether n is even or odd. The polynomial A is called the **scaling function.** Finally, set

$$A^*(t) = \int_a^t A(x)dx. \tag{1.3}$$

THEOREM 1.1. *With notation and assumptions as introduced above, we have*

$$\int_a^b f(x)dx = \int_a^b p(x)dx - \frac{f^{(2k+2)}(\xi)}{(2k+2)!}\int_a^b A^*(t)dt \tag{1.4}$$

for some ξ such that $a < \xi < b$.

The first integral on the right in (1.4) is easily evaluated, as p is a polynomial, and yields the Newton-Cotes quadrature formula for $n+1$ points. This formula approximates the integral of f in terms of the values of f at the points $x_0, x_1, \cdots, x_n$. The integral in the second expression on the right in (1.4) is independent of f (and also easily evaluated as A^* is a polynomial). This expression, therefore, gives the error of approximation in terms of the value of the $(2k+2)$-nd derivative of f at some (unknown!) point between a and b.

Theorem 1.1 for $n > 1$ is proved in section 4. Sections 2 and 3 contain the proofs of two auxiliary results. The idea of our proof of Theorem 1.1 is basically the same as that of Steffensen. We write $f(x) = p(x) + e(x)$ and attempt to estimate $\int_a^b e(x)dx$. Following Steffensen, we put $e(x) = A(x)g(x)$; that is, we define g by

$$g(x) = \frac{f(x) - p(x)}{A(x)} . \tag{1.5}$$

(This is a departure from Steffensen if n is odd.) The function g is defined wherever f is defined except at the points $x_0, x_1, \cdots, x_{2k}$. Now f has at least two derivatives since $n \geqq 1$ and we have assumed $2k+2$ derivatives. Therefore, f' is

continuous, and we can define

$$g(x_i) = \lim_{x \to x_i} \frac{f(x) - p(x)}{A(x)} = \frac{f'(x_i) - p'(x_i)}{A'(x_i)} \qquad (i = 0, 1, \cdots, 2k) \tag{1.6}$$

by l'Hospital's rule. With this definition, g is defined and continuous in an open interval containing $[a, b]$.

The proof in section 4 does not apply if $n=1$ unless f happens to have a third derivative. Our assumption of $2k+2$ derivatives only yields two derivatives for $n=1$; and as we do not wish to make unnecessary assumptions, we note that a proof of Theorem 1.1 with $n=1$ is given as an exercise with hints in Hardy [1, p. 330]. For the remainder of the paper, therefore, we may assume that $n > 1$. This assures us that our function f has at least four derivatives. In section 2 we need to know that g' exists, and in section 4 we need to know that g' is continuous. If f has four derivatives, these two results follow from l'Hospital's rule as follows: First, if x is not one of the points $x_0, \cdots, x_{2k}$, then g is differentiable by (1.5) and the quotient rule for derivatives; and one obtains

$$g'(x) = \frac{A(x)[f'(x) - p'(x)] - A'(x)[f(x) - p(x)]}{[A(x)]^2} \cdot \tag{1.7}$$

Therefore, we have to show in addition that

$$\lim_{x \to x_i} \frac{g(x) - g(x_i)}{x - x_i} = \lim_{x \to x_i} \frac{A'(x)[f(x) - p(x)] - A(x)[f'(x_i) - p'(x_i)]}{A(x)A'(x_i)(x - x_i)}$$

exists for $i=0, 1, \cdots, 2k$. This can be done by two applications of l'Hospital's rule. As $f^{(3)}$ exists, f'' is continuous, and we have

$$g'(x_i) = \frac{A'(x_i)[f''(x_i) - p''(x_i)] - A''(x_i)[f'(x_i) - p'(x_i)]}{2[A'(x_i)]^2} \tag{1.8}$$

for $i=0, 1, \cdots, 2k$. Therefore, g' exists. To prove g' continuous, we note first that this is obvious from (1.7) if x is not one of the points $x_0, \cdots, x_{2k}$. We have left, therefore, to prove that

$$\lim_{x \to x_i} g'(x) = g'(x_i) \qquad (i = 0, 1, \cdots, 2k).$$

This can be done by evaluating the limit of (1.7) as $x \to x_i$ by again two applications of l'Hospital's rule. One obtains the right hand side of (1.8). One needs the continuity of $f^{(3)}$, which of course is at hand since f has four derivatives.

2. The derivative of g. In this section, an estimate is derived for $g'(x)$ on $[a, b]$. This estimate will be needed in section 4. A similar result has been obtained by Ralston in [2]. The reader will note that the fact that the points $x_0, x_1, \cdots, x_{2k}$ are equally spaced in $[a, b]$ is never used so that we actually prove a more general theorem than is stated.

THEOREM 2.1. *For each $x \in [a, b]$, there exists a number ξ with $a<\xi<b$ such that*

$$g'(x) = \frac{f^{(2k+2)}(\xi)}{(2k+2)!}\,. \tag{2.1}$$

Proof. We use the standard technique of repeated application of Rolle's theorem. For fixed $x \in [a, b]$, define a new function F on an open interval containing $[a, b]$ as follows:

$$F(t) = f(t) - p(t) - A(t)[g(x) + (t - x)g'(x)]. \tag{2.2}$$

We shall need $F^{(2k+2)}$, which exists since f has $2k+2$ derivatives. In fact

$$F^{(2k+2)}(t) = f^{(2k+2)}(t) - (2k+2)!\,g'(x). \tag{2.3}$$

To prove this, note that the $(2k+2)$-nd derivatives of $p(t)$ and $A(t)$ both vanish since these functions are polynomials of degree less than $2k+2$; and $A(t)\cdot(t-x)$ is a monic polynomial of degree $2k+2$ and so has derivative $(2k+2)!$.

Our aim now is to show that $F''(t)$ has at least $2k+1$ zeroes in the interval $[a, b]$. First, by a simple calculation,

$$F'(t) = f'(t) - p'(t) - A(t)g'(x) - A'(t)[g(x) + (t - x)g'(x)] \tag{2.4}$$

and

$$F''(t) = f''(t) - p''(t) - 2A'(t)g'(x) - A''(t)[g(x) + (t - x)g'(x)]. \tag{2.5}$$

If one now differentiates the identity

$$g(t)A(t) = f(t) - g(t)$$

twice and then puts $t=x$, he will discover that
(A) $F'(x)=0$ and
(B) $F''(x)=0$ if x is one of the points $x_0, x_1, \cdots, x_{2k}$.

We must now consider two cases:

CASE 1: *x is not one of $x_0, x_1, \cdots, x_{2k}$.* In this case, F obviously has $2k+2$ roots in $[a, b]$, namely, x itself and the points $x_0, x_1, \cdots, x_{2k}$. By Rolle's theorem, F' will have at least $2k+1$ roots in $[a, b]$, none equal to x (since x is one of the $2k+2$ roots of F used in the application of Rolle's theorem). By (A) above, F' also has x as a root and hence has $2k+2$ distinct roots in $[a, b]$. Now Rolle's theorem applied to F' shows that F'' has at least $2k+1$ roots in $[a, b]$.

CASE 2: *x equals one of $x_0, x_1, \cdots, x_{2k}$.* In this case, F has the $2k+1$ roots $x_0, x_1, \cdots, x_{2k}$. Therefore, F' has $2k$ roots, none equal to x, in $[a, b]$. But $F'(x)=0$ by (A), so that F' has at least $2k+1$ roots in $[a, b]$. This implies that F'' has $2k$ roots, none equal to x, in $[a, b]$. But $F''(x)=0$ by (B) in this case. Therefore, F'' has $2k+1$ roots in $[a, b]$.

We can now prove that $F^{(2k+2)}$ has at least one root ξ in the *open* interval (a, b). This follows at once from the fact that F'' has $2k+1$ roots in $[a, b]$ by repeated application of Rolle's theorem. If now we substitute $t=\xi$ in (2.3), we get our result. This completes the proof of the theorem.

3. The scaling function. The fact that the integral of the scaling function satisfies

$$A^*(t) = \int_a^t A(x)dx \geqq 0 \tag{3.1}$$

for all $t \in [a, b]$ is required in the next section. Our aim in this section is to prove this result. We first introduce a change of variables which brings the integral in (3.1) to a more manageable form. Put

$$x = h(u) = \left(\frac{b-a}{n}\right)u + \left(\frac{b+a}{2}\right)$$

if n is even and

$$x = h(u) = \left(\frac{b-a}{n}\right)u + \left(\frac{ka+kb+a}{n}\right)$$

if n is odd. Then the increasing linear map $h(u)$ carries either the interval $[-k, k]$ or the interval $[-k, k+1]$ onto $[a, b]$ according as n is even or odd. Therefore, if $t \in [a, b]$, we find after a little calculation that

$$A^*(t) = \int_a^t A(x)dx = \int_{-k}^{t_1} A(h(u))\cdot h'(u)du = \left(\frac{b-a}{n}\right)^{2k+2}\int_{-k}^{t_1} A_1(u)du \tag{3.2}$$

for some

$$t_1 \in \begin{cases} [-k, k] & \text{for } n \text{ even} \\ [-k, k+1] & \text{for } n \text{ odd,} \end{cases} \tag{3.3}$$

and where

$$A_1(u) = \prod_{i=-k}^{k} (u+i). \tag{3.4}$$

Of course, $A_1(u)$ is the scaling function for the special symmetric case when $x_0 = -k$ and the points $x_0, x_1, \cdots, x_n$ are spaced one unit apart.

Now $A_1(u) \geqq 0$ for $u \geqq k$ as the factors in (3.4) are all nonnegative for these values of u. Therefore, in order to establish (3.1), it suffices by (3.2) and (3.3) to show that

$$\int_{-k}^{t_1} A_1(u)du \geqq 0$$

for every $t_1 \in [-k, k]$. The remainder of this section is devoted to a proof of this latter result. For notational convenience, we drop the subscript on t_1, as there is no longer any danger of confusion. We also put $v_r = -k+2r$.

LEMMA 3.1. *The sign of $A_1(u)$ is determined by the expression $(-1)^s$ where s is an integer, $0 \leqq s < 2k$, and $u \in (-k+s, -k+s+1)$.*

Proof. It is only necessary to note from (3.4) that for $u \in (-k+s, -k+s+1)$, there are $2k-s$ negative factors of $A_1(u)$, and thus the sign of $A_1(u)$ is $(-1)^{2k-s} = (-1)^s$. This completes the proof.

LEMMA 3.2. *For each integer r with $0 \leqq r \leqq (k-2)/2$, we have $\int_{v_r}^t A_1(u)du \geqq 0$ whenever $t \in [-k+2r, -k+2r+2]$.*

Proof. The proof is immediate from Lemma 3.1 if $t \in [-k+2r, -k+2r+1]$; hence we can assume $t \in [-k+2r+1, -k+2r+2]$. For t in this interval we have,

$$\int_{v_r}^{t} A_1(u)du = \int_{v_r}^{v_r+1} A_1(u)du + \int_{v_r+1}^{t} A_1(u)du.$$

This sum will be $\geqq 0$ if

$$\left| \int_{v_r+1}^{t} A_1(u)du \right| \leqq \int_{v_r}^{v_r+1} A_1(u)du.$$

Now,

$$\left| \int_{v_r+1}^{t} A_1(u)du \right| = \int_{v_r+1}^{t} -A_1(u)du$$

since for $u \in (-k+2r+1, -k+2r+2)$, $-A_1(u) > 0$ by Lemma 3.1. Thus, we need only show

$$\int_{v_r+1}^{t} -A_1(u)du \leqq \int_{v_r}^{v_r+1} A_1(u)du$$

for every $t \in (-k+2r+1, -k+2r+2)$. But, since $-A_1(u) \geqq 0$ for u in the above interval, it suffices to show that

$$\int_{v_r+1}^{v_r+2} -A_1(u)du \leqq \int_{v_r}^{v_r+1} A_1(u)du.$$

If we make the substitution $u = s+1$ in the left hand integral above, we are reduced to proving that

$$\int_{v_r}^{v_r+1} -A_1(s+1)ds \leqq \int_{v_r}^{v_r+1} A_1(s)ds.$$

We need only demonstrate that

$$-A_1(s+1) \leqq A_1(s)$$

for every $s \in (-k+2r, -k+2r+1)$. To prove this, form the ratio

$$\frac{-A_1(s+1)}{A_1(s)} = \frac{-\prod_{i=-k}^{k}(s+1-i)}{\prod_{i=-k}^{k}(s-i)} = \frac{s+k+1}{k-s}.$$

Since $-k+2r < s < -k+2r+1$ and $r \leqq (k-2)/2$, it follows that $s \leqq -1$. This in turn implies that $2s < -1$. Adding $-s+k+1$ to both sides of this last inequality, we find that

$$s+k+1 < k-s \quad \text{or} \quad \frac{s+k+1}{k-s} < 1$$

as $k-s > 0$. Thus, since $A_1(s) > 0$ for $s \in (-k+2r, -k+2r+1)$ by Lemma 3.1, we conclude that $-A_1(s+1) < A_1(s)$. This completes the proof of Lemma 3.2.

LEMMA 3.3. *We have* $\int_{-k}^{t} A_1(u)du \geqq 0$ *for each* $t \leqq 0$.

Proof. Put $t^* = t$ unless k is odd and $t > -1$, in which case put $t^* = -1$. Then for some $r \leqq (k-2)/2$, we have

$$\int_{-k}^{t} A_1(u)du \geqq \int_{-k}^{-k+2} A_1(u)du + \int_{-k+2}^{-k+4} A_1(u)du + \cdots + \int_{-k+2r}^{t^*} A_1(u)du,$$

where $t^* \in [-k+2r, -k+2r+2]$. If k is odd and $t > -1$, then the inequality follows as $A_1(u) \geqq 0$ for $u \in [-1, 0]$. By Lemma 3.2, each term in the above sum is nonnegative. Therefore, the same is true of the sum, and this completes the proof.

We are now in a position to establish our result.

THEOREM 3.4. *The integral* $\int_{-k}^{t} A_1(u)du$ *is nonnegative for* $-k < t < k$.

Proof. The theorem is true when $t \leqq 0$ by Lemma 3.3, so it will only be necessary to demonstrate the result for $0 < t < k$. Since $A_1(u) = u\prod_{s=1}^{k}(u^2 - s^2)$ is an odd function, we have

$$\int_{-k}^{t} A_1(u)du = \int_{-k}^{-t} A_1(u)du.$$

Thus, since $t > 0$ implies $-t < 0$, we have $\int_{-k}^{-t} A_1(v)dv \geqq 0$ by the results of Lemma 3.3. This completes the proof.

We note one final result. Equation (3.2) and the fact that A_1 is an odd function show that

$$A^*(b) = \int_a^b A(x)dx = 0 \tag{3.5}$$

for n even.

4. Proof of Theorem 1.1. We are now in a position to derive the standard Newton-Cotes quadrature formulas with error term for both an odd and even number of interpolating points.

From the definition of the essential error term $g(x)$ we have $f(x)=p(x)+A(x)g(x)$, and hence

$$\int_a^b f(x)dx = \int_a^b p(x)dx + \int_a^b [A(x)g(x)]dx. \tag{4.1}$$

The last integral of (4.1) can be integrated by parts. Thus,

$$\int_a^b [A(x)g(x)]dx = A^*(x)g(x)]_a^b - \int_a^b [A^*(x)g'(x)]dx, \tag{4.2}$$

where $A^*(t)$ is defined by (1.3). Now $A^*(b)=0$ by (3.5) if n is even; and $g(b)=0$ if n is odd from the definitions. Also, $A^*(a)=0$ by (1.3). Hence the first term of the right member of (4.2) is zero. By the results of section 3, $A^*(t)\geqq 0$ on $[a, b]$ and so we may apply the mean value theorem for integrals which gives

$$\int_a^b [A^*(x)g'(x)]dx = g'(X)\int_a^b A^*(x)dx \tag{4.3}$$

for some X, $a\leqq X\leqq b$. An estimate for $g'(X)$ is given by Theorem 2.1. Hence,

$$\int_a^b [A^*(x)g'(x)]dx = \frac{f^{(2k+2)}(\xi)}{(2k+2)!}\cdot\int_a^b A^*(x)dx,$$

where $a<\xi<b$. Substituting this result in (4.1), we have Theorem 1.1.

This paper consists in part of results appearing in the second named author's Master's Thesis at the University of Tennessee.

References

1. G. H. Hardy, A Course of Pure Mathematics, Cambridge, 10 ed., 1958.
2. A. Ralston, On differentiating error terms, this Monthly, 71 (1963) 187–189.
3. J. F. Steffensen, Interpolation, Chelsea, 2 ed., New York, 1950.

SYSTEMS OF ROADS WITH COUNTERS

ROBERT KOPP, San Diego State College

In this paper we shall try to develop some ideas which will lead to a solution of problem E 1980 which appeared on page 438 of the April 1967 edition of this Monthly and was proposed by R. E. Chandler:

E 1980: "The street plan for a certain town is a generalization of the tic-tac-toe board with n vertical streets and m horizontal streets (with no vertical street intersecting the end of a horizontal street and vice versa). An automobile

will start at the end of one (unspecified) street and travel to the end of another. What is the minimum number of traffic counters (each to be placed in the middle of a block) necessary to determine the route of the automobile? The route is to have no cycles, and the counters record the time that the automobile passes so that the direction of the route is determined when two counters have been passed."

1. Preliminary definitions and observations. Let G be a connected graph. An edge of G, no portion of which, with the exception of one end-point, meets any other edge of G will be called an **end**.

By a **route** of G we shall mean a path in G containing exactly two ends.

A graph G with the property that each edge belongs to some route will be called a **system of roads.**

In this paper we shall provide a method for determining the minimum number of counters necessary to distinguish routes on such graphs and also the location of the counters on the system of roads.

It is no loss of generality to restrict consideration to such graphs. For suppose G is an arbitrary connected graph. Remove all edges which are not part of some route to obtain a new graph G'. G' is easily seen to be connected and each of its edges is a part of some route. Find the minimum number of counters for G' and their position on G'. Then reinsert the edges deleted from G.

We shall use the word **route** in two ways, one way as in E1980, and the second way to mean the equivalence classes consisting of pairs of routes lumped together without regard to direction of travel. When this is the case we shall italicize the word and write it *route*.

DEFINITION 1. *A* ***spanning tree*** *of a connected graph is a connected subgraph containing all the vertices of the graph but no cycles.*

DEFINITION 2. *An edge of the complement of a spanning tree is called a* ***chord.***

It is known [1, p. 26] that there are $e-v+1$ chords associated with a spanning tree for a graph, where e and v are the number of edges and vertices respectively of the graph. The number $e-v+1$ is called the **nullity** of the graph. Some other names for this are the first Betti number or the cyclomatic number.

DEFINITION 3. *A system with counters is called* ***complete*** *if the arbitrary route of an automobile can be determined by inspecting the counters it has tripped.*

DEFINITION 4. *A system with k counters will be called* ***minimal*** *if*

(i) *it is complete,*

(ii) *if it is also complete with l counters, then $k \leqq l$.*

OBSERVATION 1. If a system is complete and a edge in the system has two counters on it, then without altering the completeness of the system one of those counters may either be placed on a vacant end or discarded.

To see that this is so, consider all distinct *routes* passing through the given

edge with two counters on it. Since the system is complete, all but possibly one of these *routes* must have tripped at least one counter in addition to the two on the given edge. Consequently the only need for two counters on this edge is to determine the direction along the exceptional *route*. In the event there is no exceptional *route* one of the two counters may be discarded without altering the completeness of the system.

This argument also provides a proof of our next observation.

OBSERVATION 2. If a system is minimal and an edge has two counters on it, there is one and only one *route* containing this edge and exactly two counters.

OBSERVATION 3. In a minimal system no edge has more than two counters. This follows directly from Observation 2.

2. The fundamental principle for systems for roads with counters. In order to prove the Fundamental Principle we shall need the following:

LEMMA 1. *Suppose an edge AB, not an end, in a minimal system has a counter on it. Then either there is a simple curve δ_{AB} in the system with ends at A and B and no counters or else the counter on AB is used only to distinguish direction along some unique route and so may be placed at an end.*

Proof: The counter is necessary if the system is minimal. Consequently it must be used to distinguish two routes θ and ϕ. These two routes have either the same edges throughout in which case it is easily seen by completeness all such pairs must have at least one common end, or else there is a segment in θ which is not on ϕ. In the first case the counter on AB must be used to distinguish direction of travel and so there is at most one counter in addition to the one on AB along θ and ϕ and so the counter may be placed at an end. Let us consider the second case. By completeness we know θ and ϕ must both have a counter along them besides one on AB. Since these routes are indistinguishable except by the counter at AB, any counter on θ and not on ϕ could only be on AB. Now start a route at the beginning of θ toward A and travel along until we just pass the last counter common to θ and ϕ. (θ and ϕ must cross or one of them would have no counter. We may suppose they cross before A.) Call this point X. Some place between X and A, θ and ϕ diverge since they are routes not containing the same edges and are distinguished only by a counter at AB. Call this point Y. Now continue along θ to point A. Then θ and ϕ do not cross (by completeness) on θ between Y and B since there can be no counters along here on either θ or ϕ (if so this would obviate the need of a counter on AB). After leaving B they must join once more at a point Z. (If they didn't, they would go to different ends, say θ to HK and ϕ to MN. But then the route from HK back on θ across BA to Y then forward along ϕ to MN would have only one counter on it at AB since the counter at AB is necessary to distinguish θ from ϕ.) Again by reasoning similar to that above there is no counter on the section of θ from B to Z. Consequently the path A to Y along θ, Y to Z along ϕ and Z to B along θ is the simple curve δ_{AB}.

It is an immediate consequence of the lemma, that we may arrange a minimal system (with at least two ends) if we so desire so that both

(1) Given any two nodes A and B there is a simple curve δ_{AB} with no counters joining A and B.

(2) Each end has a counter and the only counters distinguishing direction are on ends.

To see (1), simply observe that the system is connected so it is possible to get from A to B by a finite number of edges

$$AA_1,\ A_1A_2,\ \cdots,\ A_nB.$$

But by the lemma we may assume either A_kA_{k+1} has no counter on it or else we may bypass it with a curve $\delta_{A_kA_{k+1}}$ without counters. The curves $\delta_{A_kA_{k+1}}$ and $\delta_{A_{k+1}A_{k+2}}$ cannot intersect except at A_{k+1} or in an arc, because of completeness of the system. (2) is an immediate consequence of (1).

The possibility of arranging a minimal system so (1) and (2) are satisfied will be called *The Fundamental Principle of Systems of Roads with Counters*, and we shall henceforth assume that all minimal systems are so arranged. (Note: if a system is minimal and has more than two ends there *must* be a counter at every end.)

THEOREM 1. *The minimum number of counters required to distinguish the routes in a system of roads S is equal to the number of ends of the system plus the nullity of the system. If a counter is placed on each end of S and on each chord for some fixed spanning tree we obtain a minimal system.*

Proof: We know we may assume a minimal system S has a counter at each end. In addition there must be a counter on at least one edge of every cycle. Let T be the set of edges without counters. Then T is a tree of the system S and thus the number of edges with counters must be at least equal to the nullity of S. Thus the number of counters must be at least equal to the number of ends plus the nullity of S.

Conversely, let T be a spanning tree of S and put a counter on each end of S and on each chord of T (hence, number of counters = number of ends plus nullity). Then S is complete. For if these counters did not distinguish two routes R_1 and R_2 then the set of lines belonging to either R_1 or R_2 but not both contains a cycle without counters. Yet this is a subset of T which is a contradiction.

3. Solution to E1980. In Figure 1 (p. 1076) we have an $m\times n$ system and in Figure 2 a spanning tree. From this we see the nullity of the system is $(n-1)$ $(m-1)$. There are $2n+2m$ ends in the system. Consequently a minimal system must have $2n+2m+(n-1)(m-1)=(n+1)$ $(m+1)$ counters. Such a system is illustrated in Figure 3.

Acknowledgements. I should like to thank Professors Edgar Howard, Daniel Saltz, and Nick Morez for the suggestions they made. I am also most grateful to the referee for considerably simplifying this paper.

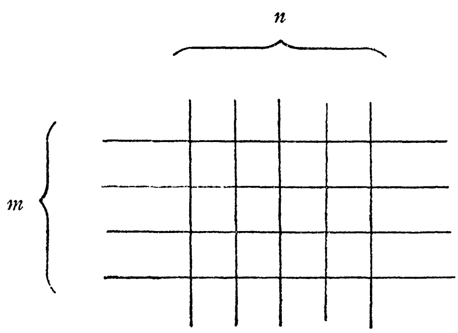


Fig. 1

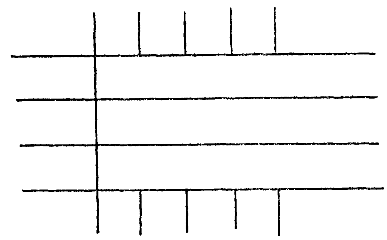


Fig. 2

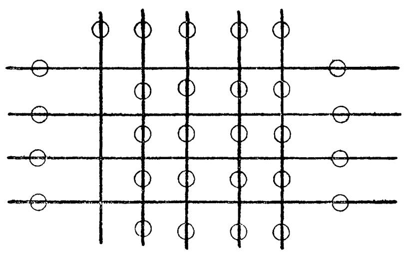


Fig. 3

Reference

1. Seshu and Reed, Linear Graphs and Electrical Networks, Addison-Wesley, Reading, Mass., 1961.

Note. Problem E 1980 was solved by others after Professor Kopp's article was accepted. See the October MONTHLY, page 883. *Editor.*

MATHEMATICAL NOTES

EDITED BY DAVID DRASIN

Manuscripts for this Department should be sent to David Drasin, Division of Mathematical Sciences, Purdue University, Lafayette, IN 47907.

COMPLEMENTS AND COMMENTS

DAVID DRASIN, Purdue University

1. The end of another volume of Mathematical and Classroom Notes prompts the annual review of correspondence based on material we have published. I wish to thank those readers who have communicated the information below.

2. Elementary mathematics (material covered normally during the first

two undergraduate years). W. C. Waterhouse has observed that R. P. Boas's proof of Lhospital's rule (Nov. 1969, p. 1051) is essentially the same as one given by F. Lettenmeyer in J. Reine Angew. Math., 174 (1935) 246.

M. V. Subbarao in the September 1968 MONTHLY, p. 772, gave an elementary demonstration of the fact that if n is a natural number, then $\sqrt{n}$ is either an integer or irrational. At the end of the paper, he asked whether the method could be extended to the kth roots, for $k>2$. In fact, I. Niven and E. A. Maier had surmounted this difficulty in a paper which appeared in MATHEMATICS MAGAZINE, 1964, pp. 208–210.

3. Algebra. Richard Singer, in the December 1969 issue, p. 1131, gave proofs of the standard theorems of Gauss and Eisenstein, using the natural morphism $Z[X]\to Z_p[X]$. That proof of Gauss's theorem also appeared in Richard A. Dean's book *Elements of Abstract Algebra.* Dean comments that Singer's proof of the irreducibility criterion is less general than the traditional one, in that the last line of Singer's proof does not go through for an arbitrary UFD. For if D is a UFD and p is a prime in D, then $D_p=D/(p)$ need not be a UFD; a common counterexample is $D=Z[X]$ and $p=x^2+5$.

Since C. E. Linderholm's article "A group epimorphism is surjective" (Feb. 1970, p. 176) appeared, the author discovered that W. Burgess had already published this fact in the Canadian Math. Bulletin, 1965, pp. 759–769.

Peter Yff calls attention to some facts about finite groups which may have been overlooked. Presumably the results are due to A. Speiser, since they appear in his book *Die Theorie der Gruppen von endlicher Ordnung,* (3rd ed., Dover, 1943). Part (i) of Theorem 1 in "A generalization of Frobenius's theorem" by H. G. Bray is covered in Speiser's book (pp. 112–113), albeit not stated as a theorem; the fact also appears on p. 62 of I. N. Herstein's text *Topics in Algebra.* The remainder of Theorem 1 is covered by the following general result which appeared on p. 117 of Speiser's book: *If H is a subgroup of index n in a finite group G, and if N is the largest normal subgroup of G contained in H, then the index of N in H is a divisor of $n!$.*

Many of the results in J. A. Eagon's paper "Finitely generated domains over Jacobson semi-simple rings are Jacobson semi-simple" were obtained earlier by R. Gilmer (Pacific Journal, 1966, pp. 275–284).

Gilmer also observes that it is possible to give many examples of irreducible polynomials which are reducible mod p for all p; M. A. Lee had given some in the December 1969 issue, p. 1125. For example (see article by W. J. Guerrier, Jan. 1968, p. 46) if n is a positive integer such that the multiplicative group of units of $Z/(n)$ is not cyclic, then the nth cyclotomic polynomial F_n is irreducible in $Z[X]$ but reducible in $(Z/(p))[X]$ for each prime p. Those positive integers n for which the multiplicative group of units of $Z/(n)$ is cyclic are known; cf. exercises 4–6 of vol. 2 of van der Waerden's *Modern Algebra* (English ed. pp. 114–115).

The note by R. Datko and V. Seshadri entitled "A characterization and a

canonical decomposition of Hurwitzian matrices," vol. 77, August–September 1970, pp. 732–733, is similar to a result of Olga Taussky Todd which appeared in the Jour. of Math. Anal. and Appl. (vol. 2, #1, Feb. 1961, pp. 105–107). Prof. Todd's result proves that every real stable matrix is unitarily similar to a matrix of the form $(-I+S)P$, where S is skew symmetric and P is a diagonal matrix with positive elements. Thus in the case of real matrices, since the matrix SP is similar to the matrix $P^{1/2}SP^{1/2}$, the results of Datko and Seshadri follow as a consequence. The extension to the complex case is routine. The method of proof used in the two papers is different, except that the starting point is a theorem of A. M. Lyapunov (see, e.g., Gantmacher, F. R., *Theory of Matrices*, vol. II).

4. Topology. P. M. Rice reports that there is a gap in the proof of Lemma 1 of his paper "A topological characterization of the real numbers" (Feb. 1969, p. 184); indeed he has a counterexample. The main result of the paper is, of course, correct; however, the short proof presented by Rice now seems in question.

Theorem 2 of N. Levine's article "Dense topologies" (Oct. 1968, pp. 847–852) is also incorrect, but for typographical reasons; the statement of the theorem should be "every dense subspace of a D-space is a D-space." The slip was pointed out to us by Sabah A. Ghullam.

Some time ago (March 1967, pp. 261–266) A. Wilansky wrote "Between T_1 and T_2." The main results of J. E. Joseph's article (Dec. 1969, p. 1125) are contained in Wilansky's note.

5. Number theory. In the May 1970 issue, pp. 510–512, G. J. Simmons showed that given r and n, there exist infinitely many integers $m \geqq r$ such that

$$\left(\binom{m}{r}, N\right) = 1. \tag{A}$$

However, it is possible to completely characterize those m for which (A) holds, using the article "On periodicities of certain sequences of residues" of W. F. Trench, which appeared in this MONTHLY, August-September 1960, pp. 652–656. Trench offers the following procedure. Define $N_0 = \prod_{i=1}^{n} p_i$, where $p_1, \cdots, p_n$ are the distinct prime factors of N. Then (A) holds if and only if

$$\left(\binom{m}{r}, N_0\right) = 1,$$

which is in turn equivalent to

$$\binom{m}{r} \equiv q \pmod{N_0}, \quad \text{where} \quad (q, N_0) = 1.$$

Lemma 1 of Trench's paper states that the sequence of residues

$$\binom{m}{r} \bmod N_0; \quad m = \cdots -1, 0, 1, \cdots$$

is periodic with fundamental period

$$P = \prod_{i=1}^{n} p_i^{s_i},$$

where $s_1, \cdots, s_n$ are integers such that $p_i^{s_i-1} \leqq r < p_i^{s_i} (1 \leqq i \leqq n)$. Therefore an integer m satisfies (A) if and only if it is of the form $m = m_0 + kP$, where k is an integer, $r \leqq m_0 \leqq P-1$ and

$$\left(\binom{m_0}{r}, p_i\right) = 1 \qquad (1 \leqq i \leqq n).$$

6. Analysis. In 1947, R. Arens (Bull. AMS, p. 623) proved that a normed algebra in which $\|xy\| = \|x\| \|y\|$ is either the reals, comlpexes, or quaternions. This generalises the celebrated Gelfand-Mazur theorem and contains as a special case the note of Howard A. Seid (March 1970, p. 282); this reference was noted by Enzo R. Gentile.

Gerald A. Heuer observes that a paper which he and some undergraduates published in the April 1965 MONTHLY, pp. 370–373, contains those results in J. E. Nyman's note "An application of diophantine approximation" (July 1969, pp. 668–671). Later, Heuer himself (April 1966, pp. 378–379) improved his earlier results.

R. P. Boas calls my attention to the article "Limits of integrals" by R. P. Agnew (Duke Math. J., March 1942, pp. 10–19). Agnew's result allows an immediate generalization of the theorem of my paper "An Application of Tauberian Methods to a Problem in Series" (Feb. 1970, pp. 152–156). For a suitable function $f(t)$, $(1 \leqq t < \infty)$, I let

$$\Psi(x) = \int_1^x f(t) dt,$$

and consider consequences of the assumption that $\{\Psi(\sigma x) - \Psi(x)\}$ tends to a limit (as $x \to \infty$) for each fixed $\sigma > 1$. Agnew shows that my deductions can be derived even under the apparently weaker hypotheses that $\{\Psi(\sigma x) - \Psi(x)\}$ tends to a limit (as $x \to \infty$) for a set of σ of positive Lebesgue measure or even for two appropriate values of σ.

A problem left open in "Nets and sequences, an example" was later resolved by the same author, W. M. Priestly, in a paper which has recently appeared in the Proceedings of the A. M. S. (Feb. 1970).

An example of a series $\sum f_n(x)$ which converges absolutely and uniformly, but, for some rearrangement, fails to converge uniformly, was given by D. J. Harris and B. L. D. Thorp in the September 1969 issue, pp. 801–2. P. Turán reports that a related result has appeared in *Matematikai Lapok Fasc.*, 1 (1953) p. 36; namely that there exist analytic functions $\phi_n(z)$ in $\{|z| < 1\}$ with the property that $\sum \phi_n(z)$ converges absolutely but not uniformly. The examples

were given by Ákos Csánszár and Cathleen Rényi. Rényi's example was particularly simple: $\phi_n(z) = \frac{1}{2}(1-z)\left[\frac{1}{2}(1+z)\right]^n$.

And finally, we record that the theorem given by Roger Horn (Jan. 1970, pp. 65–66) on the Wronskian test for linear independence has a long past. One reference, supplied by John Baxley, is "On the Wronskian tests for linear dependence" by Maxime Bôcher (Annals of Math., Ser. 2, v. 17 (1915–16), pp. 167–168). Bôcher himself references earlier papers, and indicates that Peano was aware that for a family of analytic functions, nonvanishing of the Wronskian becomes a necessary and sufficient test for linear dependence.

EVALUATION OF SOME INTEGRALS BY CONTOUR INTEGRATION

M. LUTZKY, U.S. Naval Ordnance Laboratory

In this note we evaluate some real integrals by contour integration, using a method involving the integration of certain logarithmic functions over portions of the unit circle. We shall use the fact that on the unit circle

$$\log(1+z) = \log(2\cos\tfrac{1}{2}\theta) + \tfrac{1}{2}i\theta,\ (-\pi < \theta < \pi), \tag{1}$$

and

$$\log(1-z) = \log(2\sin\tfrac{1}{2}\theta) + \tfrac{1}{2}i(\theta - \pi),\ (0 < \theta < 2\pi). \tag{2}$$

For our first case, we integrate the function $\log(1+z)/iz$ about the closed contour consisting of the real axis between 0 and 1, the imaginary axis between 0 and i, and that portion of the unit circle lying in the first quadrant. Since there are no singularities within the contour, Cauchy's theorem yields

$$i\int_0^1 \log(1+x)dx/x = \int_0^{\pi/2}\left\{\log(2\cos\tfrac{1}{2}\theta) + \tfrac{1}{2}i\theta\right\}d\theta + i\int_0^1 \log(1+iy)dy/y. \tag{3}$$

We have

$$\log(1+iy) = \tfrac{1}{2}\log(1+y^2) + i\arctan(y), \tag{4}$$

and a simple change of variable shows that

$$\int_0^1 \log(1+y^2)dy/y = \tfrac{1}{2}\int_0^1 \log(1+x)dx/x. \tag{5}$$

Using (4) and (5), we find that the imaginary part of (3) yields

$$\int_0^1 \log(1+x)dx/x = \pi^2/12, \tag{6}$$

while the real part gives the result

$$\int_0^{\pi/2} \log(2\cos\tfrac{1}{2}\theta)d\theta = \int_0^1 (\arctan y)dy/y. \tag{7}$$

Similarly, we can integrate $\log(1-z)/iz$ about the same contour, adding an indentation at $z=1$ to exclude the singular point. The integral around the indentation goes to zero as the indentation shrinks, so that Cauchy's theorem, (2), and (6) yield

$$\int_0^1 \log(1-x)dx/x = -\pi^2/6 \tag{8}$$

and

$$\int_0^{\pi/2} \log(2\sin\tfrac{1}{2}\theta)d\theta = -\int_0^1 (\arctan y)dy/y. \tag{9}$$

The results (6) and (8) are well known, although they are not usually arrived at by contour integral methods. Furthermore, upon addition of (7) and (9), another familiar result is obtained:

$$\int_0^{\pi/2} \log(2\sin\theta)d\theta = 0. \tag{10}$$

For our next example, we integrate $[\log(1-z)]^p/iz$ for $p=1$, 2, and 3 about the unit circle, indented at $z=1$. The integral around the indentation vanishes in the limit, and Cauchy's theorem gives

$$\int_0^{2\pi} [\log(2\sin\tfrac{1}{2}\theta) + \tfrac{1}{2}i(\theta-\pi)]^p\, d\theta = 0. \tag{11}$$

When $p=1$, the real part of (11) yields an integral equivalent to (10). When $p=2$, relation (10) and the imaginary part of (11) yield

$$\int_0^{2\pi} \theta\log(2\sin\tfrac{1}{2}\theta)d\theta = 0, \tag{12}$$

while the real part gives

$$\int_0^{\pi} [\log(2\sin\tfrac{1}{2}\theta)]^2\, d\theta = \pi^3/12. \tag{13}$$

When $p=3$, relation (13) and the imaginary part of (11) yield

$$\int_0^{2\pi} \theta[\log(2\sin\tfrac{1}{2}\theta)]^2\, d\theta = \pi^4/6. \tag{14}$$

Unlike the previous cases, however, the real part of (11) with $p=3$ does not yield a particularly simple result. After using (10) and (12) to simplify the resulting expression, we end up with

$$4\int_0^{2\pi} [\log(2\sin\tfrac{1}{2}\theta)]^3\, d\theta = 3\int_0^{2\pi} \theta^2\log(2\sin\tfrac{1}{2}\theta)d\theta. \tag{15}$$

Integrals of the form $\int_0^\pi [\log (2 \sin \frac{1}{2}\theta)]^p \, d\theta$ have been studied by Bowman [1] and Beumer [2] using real-variable methods, and it can be shown that they are expressible in terms of the Riemann zeta function of integral argument. For instance, if we substitute the Fourier series

$$\log(2 \sin \tfrac{1}{2}\theta) = -\cos \theta - \tfrac{1}{2} \cos 2\theta - \cdots \tag{16}$$

into the right hand side of (15), we obtain

$$2 \int_0^\pi [\log(2 \sin \tfrac{1}{2}\theta)]^3 \, d\theta = -3\pi \sum_1^\infty k^{-3} = -3\pi\zeta(3). \tag{17}$$

To introduce our last example, we point out that the integrands of (14) and (12) have the form $\theta f(\cos\theta, \sin\theta)$. Clearly, the efficacy of the method in dealing with integrals of this type arises from the fact that the imaginary parts of (1) and (2) are linear functions of θ. For our final example, we study another integral of this kind in which, however, the function $f(\cos\theta, \sin\theta)$ is not logarithmic. We define $g(z) = a(z-1)(z+1)/[i(z-a)(az-1)]$ for $-1 < a < 1$, and integrate $\log(1+z)g(z)/z$ about the unit circle, indented to exclude the point $z = -1$. The only contribution to the integral comes from the simple pole at $z = a$, where the residue is $-i \log(1+a)$; consequently Cauchy's theorem yields, using (1):

$$\int_{-\pi}^{\pi} [\log(2 \cos \tfrac{1}{2}\theta) + \tfrac{1}{2}i\theta] g(e^{i\theta}) d\theta = -2\pi i \log(1 + a). \tag{18}$$

It is easily shown that $g(z)$ is real-valued on the unit circle; specifically,

$$g(e^{i\theta}) = -\frac{2a \sin \theta}{a^2 - 2a \cos \theta + 1} \cdot \tag{19}$$

Therefore, the imaginary part of (18) gives the integral

$$\int_{-\pi}^{\pi} \frac{\theta \sin \theta}{a^2 - 2a \cos \theta + 1} \, d\theta = \frac{2\pi}{a} \log(1 + a). \tag{20}$$

This integral previously was treated by complex integration by Lindelöf [3], who integrated the function $z/(a - e^{-iz})$ around the rectangular contour with vertices at $-\pi$, π, $\pi + iR$, and $\pi - iR$. His method is somewhat more cumbersome than the one presented here, involving as it does the evaluation of subsidiary integrals along the sides of the contour, as well as requiring the passage to the limit $R \to \infty$.

References

1. F. Bowman, Notes on the integral $\int_0^{\pi/2} (\log \sin\theta)^n d\theta$, J. London Math Soc. 22 (1947), 172–173.

2. M. G. Beumer, Some special integrals, this MONTHLY 68(1961), 645–647.

3. E. Lindelöf, Le Calcul des Résidus, Gauthier-Villars, Paris, 1905, esp. pp 48–49.

EXISTENCE OF FOUR CONCURRENT NORMALS TO A SMOOTH CLOSED CURVE

NARSINGH DEO, California Institute of Technology, and M. S. KLAMKIN, Ford Motor Company

In a rather elegant lecture on calculus, Guggenheimer [1] establishes, among other related theorems, the existence of a point from which at least four normals can be drawn to an oval. Two such points are the mass centroid and the curvature centroid. Chakerian and Stein [2] have shown that another four normal point is the perimeter centroid. The latter proof is by a simple continuity argument whereas the proofs given by Guggenheimer cannot (at least up to now) be obtained the same way. In this note, we give a simpler intuitive geometric proof for the existence of a four normal point which can be used in elementary courses dealing with continuity methods in convexity or geometry (e.g., a la Yaglom and Boltyanskii [3]).

For an oval (smooth, convex curve), it is well known that there exists at least one largest chord and, furthermore, this chord is perpendicular to the curve at each end.

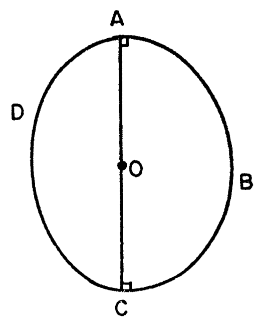


FIG. 1

We now show that the midpoint O of a largest chord AB has the desired property. If ABC in Fig. 1 is circular, our proof is over. If ABC is not circular, then there must exist points whose distance from O is either larger or smaller than AO. The point (or points) P which maximize or minimize the radial distance OP will be such that $\overline{OP}$ is normal to the curve. Geometrically, if we consider the family of concentric circles centered at O, there will exist at least one circle which is tangent to open arc ABC. Since a similar argument applies to the other open arc CDA, our proof is completed.

By a similar argument, one can show that another four normal point is the midpoint of a chord of minimal width.

For the case of C^2 closed nonconvex curves, a largest chord which is perpendicular to the curve at each end still exists. Consequently, the midpoint O will still be a four normal point provided that it falls inside the curve. If O falls on the boundary, then possibly only three distinct normals can be drawn from O. However, there are then at least four points on the curve such that the nor-

mals at these points are concurrent (note that if the curve touches itself at O, this point is counted twice). If O falls outside the curve, it is possible that two of the normals are on the same ray. Whether or not these two normals are to be considered distinct is a matter of definition.

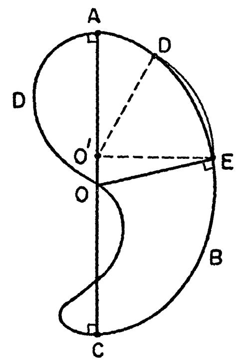


FIG. 2

For the case (Fig. 2) when O falls on the curve and the curve is also simple and contains no circular or straight segments, then for points O' of AC, sufficiently close to O, there exists a normal from O' to the open arc ABC. For the case illustrated, being sufficiently close means that if $\overline{OE}$ is normal to the curve and $OE<OA$; then $O'E$ must be $<O'A$. If $OE>OA$, then $O'E$ must be $>O'C$ (and in this case the extension of circular arc DE will intersect the open arc EBC). Also, there exists a normal from O' to the portion of the arc ADC contained in a circle of center O' and radius $OO'+\epsilon$ but here O' must satisfy $OO'<OA/2$. Hence, we have a segment of four normal points. The same also applies for an oval not containing any circular arcs. It now follows easily that the circle is the only oval with a unique "four-normal" point. Note that if the midpoint O of the maximum length chord AC turns out to be the only point on AC which is a four-normal point (e.g., if ABC was a semi-circle), then we would consider points on the minimum width chord.

Acknowledgment: The authors are grateful to Professor Guggenheimer for his suggestions improving the style of the original note as well as extending its scope.

References

1. H. H. Guggenheimer, Geometrical applications of integral calculus, p. 84 (contained in K. O. May, Lectures on Calculus, Holden-Day, San Francisco, 1967). Guggenheimer notes that since the problems treated in his paper are all of recent origin, it seemed possible to give the names of the first discoverers and a fairly complete bibliography.

2. G. D. Chakerian and S. K. Stein, On the centroid of a homogeneous wire, Mich. Math. J., 11 (1964) 189–192.

3. I. M. Yaglom and V. G. Boltyanskii, Convex Figures, Holt, Rinehart and Winston, New York, 1961.

A COMBINATORIAL LEMMA INVOLVING A DIVERGENCE CRITERION FOR SERIES OF POSITIVE TERMS

K. A. Post, Technological University Eindhoven, The Netherlands

In this paper we shall give sufficient conditions for a mapping f of the set of positive integers into itself in order to ensure that a series $\sum_{n=1}^{\infty} a_n$ diverges whenever $0 < a_n < a_{n+1} + a_{f(n)}$ for all n.

This problem came up around 1964 when Graham and Jewett showed that $f(n) = n(n+1)$ $(n = 1, 2, 3, \cdots)$ has the desired property. They were also able to show roughly that a sufficiently regular function f that grows "as fast as" $\frac{1}{2}n^2$ also must have this property. This result is not published since the method they used was rather complicated (private communication). Later on, Erdös pointed out that the condition $f(n) \leqq n^2/(2+\epsilon)$ (ϵ being a fixed positive number) is not sufficient for divergence, by taking $a_n = 1/n \log n(\log \log n)^{1+\delta}$ for sufficient small $\delta = \delta(\epsilon) > 0$. Around 1965 D. J. Newman showed that the condition $f(n) = n^2$ $(n = 1, 2, 3, \cdots)$ implies divergence.

We shall prove the following theorem:

THEOREM. *Let f be a mapping of the set of positive integers into itself satisfying the difference condition*

$$\text{(D)} \qquad f(n+1) - f(n) \geqq n + 1 \qquad (n = 1, 2, 3, \cdots).$$

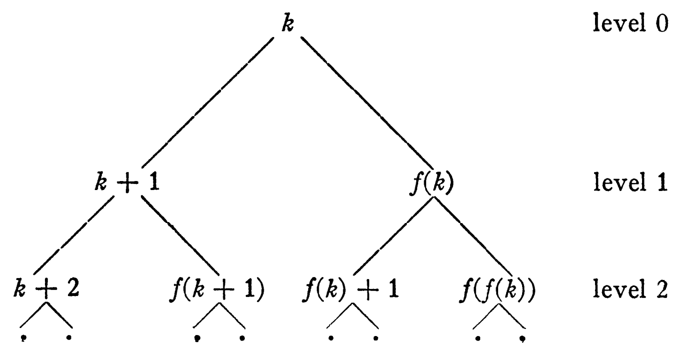
Then any series $\sum_{n=1}^{\infty} a_n$, for which

$$\text{(1)} \qquad 0 \leqq a_n \leqq a_{n+1} + a_{f(n)} \qquad (n = 1, 2, 3, \cdots)$$

is divergent or $a_n = 0$ for all $n \geqq 2$.

The proof of this theorem is based essentially upon the following lemma:

LEMMA. *Let f satisfy the difference condition (D), and let k be an integer $\geqq 3$. Consider the tree*



Then for every $t \geqq 1$ all of the 2^t numbers at level t are distinct.

Proof of the Lemma. From the difference condition (D) on f it follows that for every n

$$f(n) \geqq \sum_{j=1}^{n} j = \tfrac{1}{2}n(n+1). \tag{2}$$

Hence, $f(k) > k+1$, so that level 1 contains two distinct numbers. Now let t be the smallest number such that level t contains a number, q say, twice. Then level $t-1$ contains two numbers, r and s say, such that

$$r + 1 = f(s) = q. \tag{3}$$

Now we trace back in the tree from r up to the root k. Each element in this path is obtained from the preceding one by subtraction of unity or by taking inverse f-image. Since by (D) we have $f(s)-f(s-1) \geqq s$, each of the first $s-1$ steps in this back-tracing process must be subtraction of unity. On the other hand, the back-tracing process starting from the element s on level $t-1$ must lead to the root k in at most $s-k$ steps, so that r is actually deduced from k by successive addition of unity in the tree. In other words,

$$r = k + t - 1. \tag{4}$$

We also have $s \geqq k+t-1$, so that by (2)

$$f(s) \geqq \tfrac{1}{2}(k+t-1)(k+t) > k + t = r + 1$$

which contradicts (3); hence, the lemma is proved.

Proof of the theorem. Suppose $\{a_n\}_{n=2}^{\infty}$ is not identically zero. Let a_j be the first element $\neq 0$ in the sequence $\{a_n\}_{n=2}^{\infty}$. Then $\{a_n\}_{n=1}^{\infty}$ contains an element $a_k > 0$ $(k \geqq 3)$. Repeated application of (1) implies

$$0 < a_k \leqq a_{k+1} + a_{f(k)} \leqq \cdots \leqq \sum_{m \in \text{level } t} a_m,$$

where the sum is taken over all indices m occurring in level t of the tree with root k. (The notation is defined in the lemma.) The smallest index in level t equals $k+t$ so that apparently a_k may be majorized by a finite sum of distinct a_m with indices m greater than an arbitrary chosen number N. For example,

$$0 < a_k \leqq a_{k_1} + a_{k_2} + \cdots + a_{k_l} \qquad (N < k_1 < k_2 < \cdots < k_l).$$

Now we repeat the same process with any of the numbers $a_{k_1}, \cdots, a_{k_l}$ obtaining $a_{p_1}, \cdots, a_{p_q}$ $(k_l < p_1 < p_2 < \cdots < p_q)$, such that

$$\sum_{i=1}^{l} a_{k_i} < \sum_{j=1}^{q} a_{p_j},$$

and so on.

Finally, we get infinitely many disjoint blocks of $\{a_n\}_{n=1}^{\infty}$ each having a sum greater than a_k, so that $\sum_{n=1}^{\infty} a_n = \infty$.

REMARKS. (a) With a slight modification of the proofs we may show that a weaker version of the difference condition, viz.,

$$f(n+1) - f(n) \geqq n - C \qquad (n \geqq N)$$

(N and C being fixed) is already sufficient for divergence of $\sum_{n=1}^{\infty} a_n$.

(b) It should be noticed that for any increasing function g there exists a function $f \geqq g$ and a nonvanishing convergent series $\sum_{n=1}^{\infty} a_n$ such that $0 \leqq a_n \leqq a_{n+1} + a_{f(n)}$ for all n. To illustrate this we take a monotonic sequence $\{n_i\}_{i=1}^{\infty}$ that tends to infinity sufficiently rapidly, and define

$$a_n = \begin{cases} 2^{-i} & \text{if } n = n_i \qquad (i = 1, 2, 3, \cdots) \\ 2^{-i-1} & \text{if } n = n_i + 1 \qquad (i = 1, 2, 3, \cdots) \\ 0 & \text{otherwise.} \end{cases}$$

Finally, we take $f(n) = n_{i+1}$ $(n_{i-1}+1 < n \leqq n_i + 1)$. So we must conclude that a difference condition like (D) seems to play the central role in the theorem.

A NOTE ON HADAMARD PRODUCTS

T. E. DAVIS, Kansas State Teachers College

The *Hadamard product* of two $n \times n$ matrices $A = \|a_{ij}\|$ and $B = \|b_{ij}\|$ is $A \circ B = \|a_{ij} b_{ij}\|$. The following result is given in [2] and generalized to normal matrices in [1]. We give here a new proof based on tensor products.

THEOREM. *Let A and B be positive semi-definite Hermitian matrices with characteristic roots $\alpha_1 \leqq \alpha_2 \leqq \cdots \leqq \alpha_n$ and $\beta_1 \leqq \beta_2 \leqq \cdots \leqq \beta_n$ resp. Then $A \circ B$ is also positive semi-definite Hermitian, and each of its characteristic roots λ satisfies*

$$\alpha_1 \beta_1 \leqq \lambda \leqq \alpha_n \beta_n.$$

Proof. The tensor product $A \otimes B = \|a_{ik} B_{jl}\|$ (with row index (i, j) and column index (k, l)) is Hermitian with characteristic roots all n^2 products $\alpha_i \beta_j$. Since $\alpha_1 \geqq 0$ and $\beta_1 \geqq 0$, the smallest of these is $\alpha_1 \beta_1$ and the largest is $\alpha_n \beta_n$. Now $A \circ B$ is a principal submatrix of $A \otimes B$, so it is (obviously) also Hermitian and (by the minimax principle) its characteristic roots λ lie in the interval $\alpha_1 \beta_1 \leqq \lambda \leqq \alpha_n \beta_n$.

The generalization in [1] to normal matrices also follows this easy path once you show that each characteristic root of a principal submatrix of a normal matrix A lies in the convex hull of the roots of A.

References

1. M. Marcus and R. C. Thompson, The field of values of the Hadamard product, Archiv der Math., 14 (1963) 283–288.

2. G. Pólya and G. Szegö, Aufgaben und Lehrsätze aus der Analysis, II, Dover, New York, 1945, p. 106, #35.

RESEARCH PROBLEMS

EDITED BY VICTOR KLEE

In this Department the Monthly presents easily stated research problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Matcrial should be sent to Victor Klee, Department of Mathematics, University of Washington, Seattle, WA 98105.

A PROBLEM IN GRAPH COLORING

BRANKO GRÜNBAUM, Michigan State University
(Presently: Univ. of Washington, Seattle)

A graph is said to be *colorable by k colors* if its set of nodes can be partitioned into k disjoint subsets such that no edges connect nodes in the same subset, and is said to be *k-chromatic* if it is colorable by k colors but not by $k-1$ colors. In recent years considerable attention has been devoted to the following problem: Given positive integers k and n, with $k \geqq 3$ and $n \geqq 3$, do there exist graphs which are k-chromatic while containing no circuits of length $\leqq n$?

The following results have been obtained: Zykov [1949] and Mycielski [1955] proved the existence of k-chromatic graphs of arbitrarily large k having no circuits of length $\leqq n=3$. Descartes [1954] and Kelly-Kelly [1954] obtained the same result for $n=5$, and Nešetřil [1966] for $n=7$. Erdös [1959] established by probabilistic methods the existence for arbitrary k and n, while Lovász [1968] gave direct constructions for such graphs. The graphs constructed in all those papers are "very large" and, in particular, contain nodes of valence large in comparison to k.

It is of some interest to note that the existence of k-chromatic graphs with no circuits of length $\leqq n$ was recently (Taylor [1969]) used to answer in the negative a question in model theory posed by Mycielski [1964].

On the other hand, a result of Brooks [1941] asserts: *If all nodes of a connected graph G have valence at most k, where $k \geqq 3$, then either G is k-colorable or else G is the complete graph with $k+1$ nodes.*

Some years ago (see Erdös [1964]) observations like the above led me to formulate the following conjecture:

CONJECTURE. *If $k \geqq 3$ and $n \geqq 3$ are integers, there exist k-chromatic, k-valent graphs $G(k, n)$ that contain no circuits of length $\leqq n$.*

The progress in settling the conjecture seems to have been very slow. Graphs $G(3, n)$ are easy to construct, for example from circuits with an odd number of edges and from 3-valent graphs with girth $\geqq n$ (concerning such graphs see, for example, Sachs [1964]). Recently Chvátal [1970] constructed a 12-node graph $G(4, 3)$. In the following lines I shall describe a 25-node graph $G=G(4, 4)$ I found in 1963. As far as I know, no other cases of the conjecture have been decided.

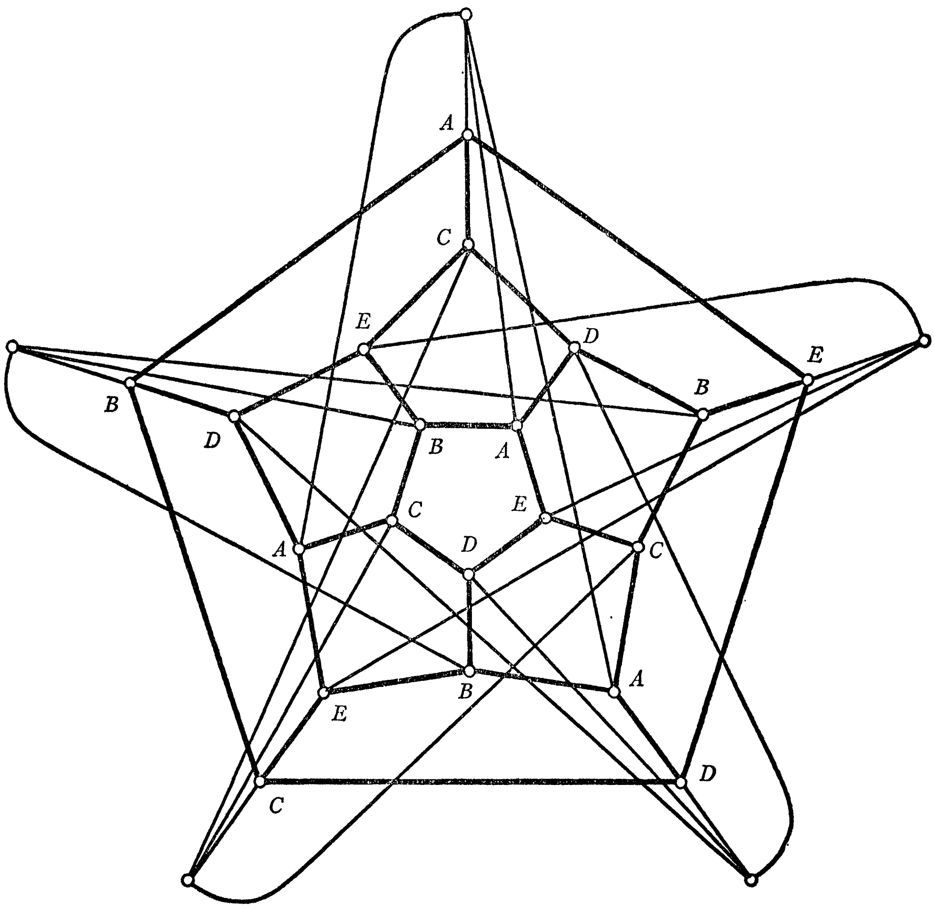


FIG. 1

The graph G (see Figure 1) may be described as follows: Out of the 25 nodes of G, 20 nodes and the edges connecting them form the net of a regular dodecahedron (heavy edges in Figure 1). It is well known that there are 10 quadruples of vertices of the dodecahedron which are the vertices of a regular tetrahedron. These ten quadruples fall into two sets of five quadruples each, the quadruples of each set covering simply the vertices of the dodecahedron. One such set is represented in the heavily drawn part of G by nodes bearing the same designation. The graph G is obtained by taking five additional nodes (the "outer" ones in Figure 1) and joining each of them to the four nodes in one of the quadruples. Thus constructed, G is obviously 4-valent, and contains no circuit with less than 5 edges. It is also easily checked that the action of the group of automorphisms of G on the five outer nodes coincides with that of the alternating group.

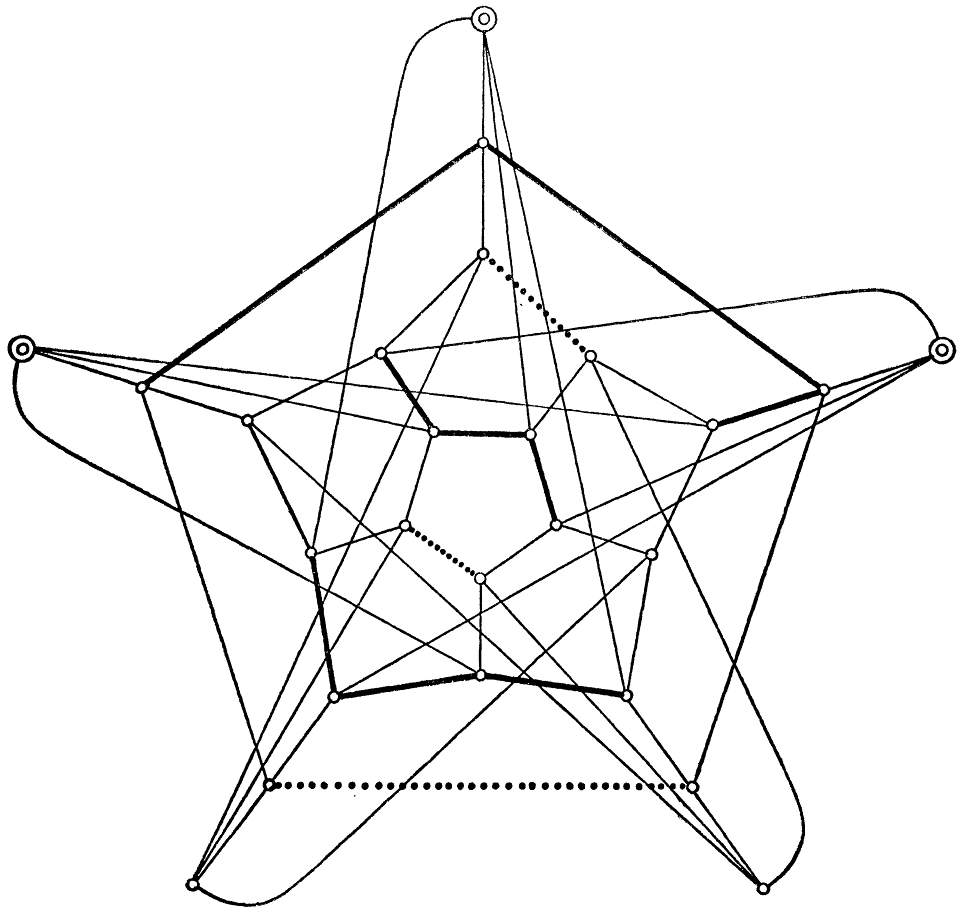


FIG. 2

In order to show that G is not 3-colorable, assume that a 3-coloring of G is given; then there are two possibilities to be considered for the outer five nodes:

(i) some three of them have the same color;

(ii) two nodes have one color, two another, the last node having the third color.

In case (i) we may, without loss of generality, assume that the three nodes in question are those marked by two circles in Figure 2; then the nodes in the paths indicated by heavy edges are colored by the remaining two colors. But then for at least one of the edges indicated by dotted lines, both nodes incident to the edge must have the same color, that of the outer three nodes.

In case (ii) we may assume, again without loss of generality, that the outer nodes are colored by 1, 2, 3 as indicated in Figure 3. Then, if the node indicated by two circles is colored 2, we arrive at a contradiction by following the arrow

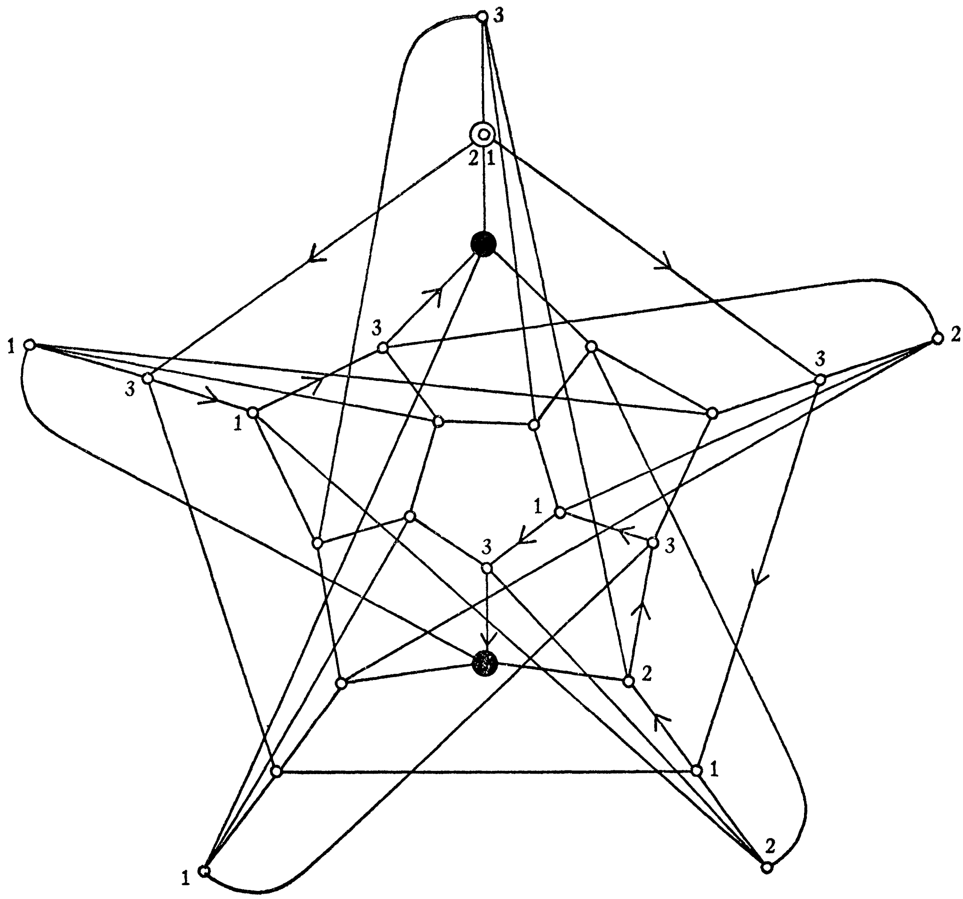


FIG. 3

leading from it to the left, there being no color available for the node indicated by the black disc. If, on the other hand, the two-circle node is colored 1 we reach a contradiction by following the arrow to the right.

Thus G is not 3-colorable, and we established that it is a graph of type $G(4, 4)$.

The author is indebted to Professor R. A. Duke for many helpful suggestions.

References

R. L. Brooks, On coloring the nodes of a network. Proc. Cambridge Philos. Soc., 37(1941) 194–197.

V. Chvátal, The smallest triangle-free-4-chromatic 4-regular graph, J. Combinat. Theory, 9 (1970) 93–94.

B. Descartes, Solution of Problem 4526, this MONTHLY, 61(1954) 352–353.

P. Erdös, Graph theory and probability, Canad. J. Math., 11(1959) 34–38.

———, Some applications of probability to graph theory and combinatorial problems. Theory of graphs and its applications, Czechoslovak Acad. Sci., Prague, 1964, pp. 133–136.

J. B. Kelly and L. M. Kelly, Paths and circuits in critical graphs. Amer. J. Math., 76(1954) 786–792.

L. Lovász, On chromatic number of finite set-systems. Acta Math. Acad. Sci. Hungar., 19(1968), 59–67.

J. Mycielski, Sur le coloriage des graphes, Colloq. Math., 3(1955) 161–162.

———, Some compactifications of general algebras, Colloq. Math., 13(1964) 1–9.

J. Nešetřil, K-chromatic graphs without cycles of length ≤ 7. (In Russian.) Comment. Math. Univ. Carolinae, 7(1966) 373–376.

H. Sachs, On regular graphs with given girth. Theory of graphs and its applications, Czechoslovak Acad. Sci., Prague 1964, pp. 91–97.

W. Taylor, Atomic compactness and graph theory, Fund. Math., 65(1969) 139–145.

A. A. Zykov, On some properties of linear complexes. (In Russian.) Matem. Sbornik, 24(66) (1949) 163–187.

CLASSROOM NOTES

EDITED BY DAVID DRASIN

Manuscripts for this Department should be sent to David Drasin, Division of Mathematical Sciences, Purdue University, Lafayette, IN 47907.

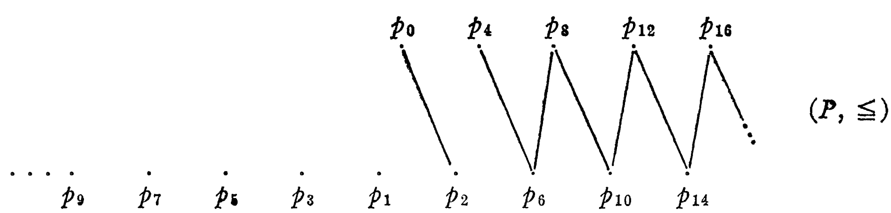
ON THE SIMILARITY OF PARTIALLY ORDERED SETS

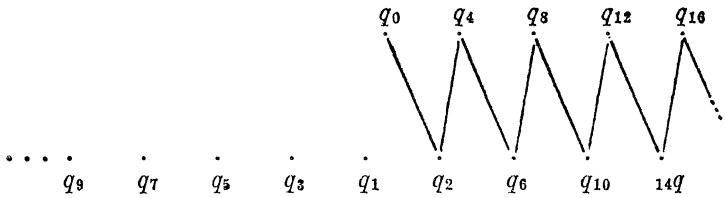
ALEXANDER ABIAN, Iowa State University

It is well known and easy to prove [1] that for linearly ordered sets $(R, \leqq)$ and $(S, \leqq')$, if there exists a one-to-one mapping p from R onto S which preserves order (i.e., $x \leqq y$ implies $p(x) \leqq' p(y)$), then $(R, \leqq)$ is similar to $(S, \leqq')$ since the converse mapping p_{-1} of p also preserves order.

It is a curious fact that in the case of partially ordered sets $(P, \leqq)$ and $(Q, \leqq')$, even a stronger hypothesis does not imply that $(P, \leqq)$ is similar to $(Q, \leqq')$. For instance, as shown in the example below (suggested by Daryl R. Fischer), if there exists a one-to-one mapping f from P *onto* Q such that f preserves order and if there exists a one-to-one mapping g from Q *onto* P such that g also preserves order, then it is not necessary that P be similar to Q, i.e., it is not necessary that there exist a one-to-one mapping h from P onto Q such that both h and its converse h_{-1} preserve order.

Example 1. Let $(P, \leqq)$ and $(Q, \leqq')$ be partially ordered sets represented in the diagram below, where $p_i < p_j$ if and only if there exists an upward connecting line from p_i to p_j. Similarly, $q_i <' q_j$ if and only if there exists an upward connecting line from q_i to q_j. Otherwise, the distinct elements are not comparable.





$(Q, \leqq')$

Now, let f be a mapping from $(P, \leqq)$ into $(Q, \leqq')$ defined by:

$$f(p_i) = q_i \quad \text{for } i = 0, 1, 2, 3, 4, \cdots.$$

Clearly, f is one-to-one onto and order preserving.

Next, let g be a mapping from $(Q, \leqq')$ into $(P, \leqq)$ defined by:

$$g(q_1) = p_2 \qquad \text{and} \qquad g(q_3) = p_0$$

and

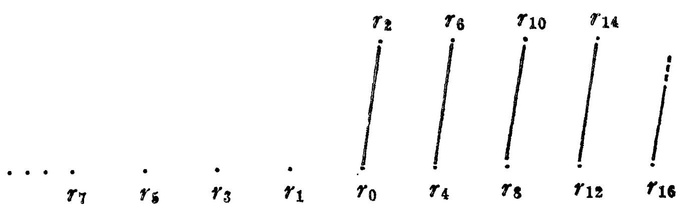
$$g(q_i) = p_{i+4} \quad \text{for } i = 0, 2, 4, 5, 6, \cdots.$$

Clearly, g is one-to-one onto and order preserving.

However, P is not similar to Q since in $(P, \leqq)$ there exists an element, namely p_2 which has one and only one successor, namely, p_0, whereas in $(Q, \leqq')$ there exists no element with one and only one successor.

Again it is a curious fact that if h is a one-to-one order preserving mapping from a partially ordered set $(R, \leqq)$ onto itself then the converse h_{-1} of h is not necessarily order preserving, i.e., h is not necessarily a similarity mapping. This is shown by the example below.

Example 2. Let $(R, \leqq)$ be a partially ordered set represented in the diagram below where $r_i < r_j$ if and only if there exists an upward connecting line from r_i to r_j. Otherwise, the distinct elements are not comparable.



$(R, \leqq)$

Now, let h be a mapping from $(R, \leqq)$ into itself defined by:

$$h(r_1) = r_0 \quad \text{and} \quad h(r_3) = r_2$$

and

$$h(r_i) = r_{i+4} \quad \text{for } i = 0, 2, 4, 5, 6, \cdots.$$

Clearly, h is one-to-one onto and order preserving.

However, h_{-1} is not order preserving since $r_0 < r_2$ whereas $h_{-1}(r_0) = r_1$ and

$h_{-1}(r_2)=r_3$ and r_1 and r_3 are not comparable. Thus, indeed h is not a similarity mapping.

In contrast to the above two examples, we have the following two theorems for the case of finite partially ordered sets.

THEOREM 1. *Let $(R, \leqq)$ be a finite partially ordered set and let h be a one-to-one order preserving mapping from R into itself. Then h is a similarity mapping from R onto itself.*

Proof. Since h is one-to-one and R is finite it is clear that h is onto. Thus, to prove the theorem it is enough to show that if $h(x)<h(y)$ then $x<y$. Let S be the set of all the ordered pairs (u, v) with $u\in R$ and $v\in R$ such that $u<v$, i.e.,

$$S = \{(u, v) \mid u \in R \quad and \quad v \in R \quad and \quad u < v\}. \tag{1}$$

Consider the mapping h^* from S into S defined by:

$$h^*(u, v) = (h(u), h(v)).$$

Since h is one-to-one, obviously, h^* is one-to-one, and since S is finite we see that h^* is one-to-one from S onto S. Now, if $h(x)<h(y)$ then by (1) we have $(h(x), h(y))\in S$. However, since h^* is onto, for some $(u, v)\in S$ we must have:

$$(h(x), h(y)) = h^*(u, v) = (h(u), h(v)). \tag{2}$$

However, since h is one-to-one we see that $x=u$ and $y=v$. But then from (2) and (1) it follows that $u<v$ and hence $x<y$, as desired.

THEOREM 2. *Let P and Q be finite partially ordered sets and let f be a one-to-one order preserving mapping from P into Q and g be a one-to-one order preserving mapping from Q into P. Then P is similar to Q.*

Proof. Since P and Q are finite and f and g are one-to-one, we see that f as well as g is an onto mapping. Clearly, to prove that P is similar to Q it is enough to show that the converse f_{-1} of f is order preserving. However,

$$f_{-1} = f_{-1}\cdot(g_{-1}\cdot g) = (f_{-1}\cdot g_{-1})\cdot g = (g\cdot f)_{-1}\cdot g. \tag{3}$$

But since f and g are one-to-one onto and order preserving, we see that the composite $g\cdot f$ is a one-to-one order preserving mapping from P onto P. But then from Theorem 1, it follows that $g\cdot f$ is a similarity mapping from P onto itself and consequently its converse $(g\cdot f)_{-1}$ is order preserving. On the other hand, since g is also order preserving we see that the composite $(g\cdot f)_{-1}\cdot g$ is order preserving. But then, from (3) it follows that f_{-1} is order preserving, as desired.

The author thanks Daryl R. Fischer for the short proof of Theorem 1.

Reference

1. W. Sierpinski, Cardinal and ordinal numbers, Warsaw, 1965, page 200.

ON CHARACTERIZATION OF UNIFORMIZABLE SPACES

J. V. MICHALOWICZ, The Catholic University of America

In most topology books which mention uniform spaces, uniformizability is characterized in terms of complete regularity. More precisely, let us say that a space X is T_{3a} if it has the property that for each $x\in X$ and closed A not containing x, there is a continuous map $\varphi: X\to I$ which has $\varphi(x)=1$ and vanishes on A; we call X completely regular if it is both T_{3a} and Hausdorff. Then we have:

THEOREM 1. *A topological space is a separated uniformizable space iff it is completely regular.*

THEOREM 2. *A topological space is uniformizable iff it is T_{3a}.*

An exception is Dugundji [1], which gives an introduction to uniform spaces that is very brief but worth building on since it presents uniform structures both in terms of entourages and uniformizing families and clearly shows their equivalence. It is the purpose of this note to show how the above theorems can easily be appended to that treatment.

The usual first example of a separated uniform space is provided by a metric space or, more generally, a gauge space (that is, a space whose topology is induced by a separating family of gauges). The main result on uniform spaces in [1] is that every separated uniformizable space is a gauge space, which yields Theorem 1 since the gauge spaces are precisely the completely regular spaces.

We show that with the addition of one definition and one simple proposition we can likewise infer Theorem 2. We define the "semi gauge space" (for want of a better term) exactly like the gauge space but with the requirement that the family of gauges be separating omitted. By the same development as before, we deduce that the uniformizable spaces are precisely the semi gauge spaces and that any semi gauge space is T_{3a}. The proof in [1] for the converse of the second statement in the completely regular case relies on embedding a completely regular space in a parallelotope, and cannot be used for the T_{3a} case. However, the required converse can also be proved in a manner that is in fact applicable to the T_{3a} case, for we have:

PROPOSITION. *Every T_{3a} space is a semi gauge space.*

Proof. Let Y be a T_{3a} space and $\mathfrak{J}$ its topology, and I^y the set of all continuous maps $f: Y\to I$. For each $f\in I^y$, $d_f: Y\times Y\to E^1$ defined by $d_f(x, y)=|f(x)-f(y)|$ is a gauge in Y. Then $\mathfrak{D}=\{d_f|f\in I^y\}$ is a family of gauges which induces $\mathfrak{J}$ since the family of all balls $B(y; d_f, \epsilon)=\{x\in Y|\,d_f(x, y)=|f(x)-f(y)|<\epsilon\}$ forms a basis for $\mathfrak{J}$. Indeed, each $B(y; d_f, \epsilon)=f^{-1}(]f(y)-\epsilon, f(y)+\epsilon[)$ is open in $\mathfrak{J}$ since f is continuous. And if $x\in U\subset Y$, where U is open in $\mathfrak{J}$ then the T_{3a} property gives a $\varphi\in I^y$ with $\varphi(x)=0$ and $\varphi(y)=1$ for $y\in \complement U$ so that $x\in B(x; d_\varphi, 1)\subset U$. Thus Y is a semi gauge space.

Reference

1. J. Dugundji, Topology, Allyn and Bacon, Boston, 1966.

AFTER THE DELUGE

D. A. MORAN, Michigan State University

The purpose of this note is to provide a somewhat simpler proof than that given by Professor Marston Morse of his elementary theorem about pits, peaks, and passes on the sphere [1]. It should be recalled that Professor Morse views a positive, real-valued, bounded, differentiable function on the sphere as an altitude, measured from the center of some hypothetical spherical planet. Critical points of the function then correspond to pits (=minima), peaks (=maxima), and passes (=saddle points of index -1) on the planet. It is assumed that no critical points more complicated than these three types ever occur, and that no two of these singularities occur at precisely the same altitude. Our viewpoint is essentially the same as this, but we start with the following slightly different

ADDITIONAL HYPOTHESIS: *No pass is as high as a peak, or as low as a pit.*

This hypothesis is easily fulfilled, if we agree to drill deep holes at the bottom of each pit, and raise tall flagpoles atop each peak. Let N_0, N_1, and N_2 be the number of pits, passes, and peaks.

Now let rain begin to fall on the planet which represents the sphere. Immediately N_0 lakes are created. As the water level rises to the altitude of a pass, a lake can merge with itself, creating an island, or else a lake can merge with another lake, resulting in a net decrease by 1 in the number of lakes. When the water level has risen to inundate every pass, but is not yet as high as any peak, the number of islands will be

$$1+\text{number of island-increasing passes},$$

and the number of lakes will be

$$N_0-\text{number of lake-decreasing passes}.$$

The number of lakes less the number of islands is therefore

$$N_0 - N_1 - 1.$$

On the other hand, at this point in time there is clearly one lake and N_2 islands, so

$$N_0 - N_1 - 1 = 1 - N_2,$$

proving Morse's generalization of the theorem of Euler.

Written while the author was partially supported by NSF grant GP 8962.

References

1. Marston Morse, *Pits, peaks, and passes* (motion picture film), Modern Learning Aids #3462, New York.

2. George Polya, Induction and Analogy in Mathematics, Princeton Univ. Press, 1954, pp. 163-165.

In conclusion, we have found the seminars to be an asset to our departmental objectives and felt that other Universities might be interested in implementing a similar program.

PROBLEMS AND SOLUTIONS

EDITED BY EMORY P. STARKE

All problems (both elementary and advanced) proposed for inclusion in this Department should be sent to E. P. Starke, 1000 Kensington Ave., Plainfield, NJ 07060. Proposers of problems are urged to enclose any solutions or information that will assist the editors. Ordinarily, problems in well-known textbooks and results in generally accessible sources are not appropriate for this Department. No solutions (except those accompanying proposals) should be sent to Professor Starke.

ELEMENTARY PROBLEMS

Solutions of Elementary Problems should be sent to Problems Group, Mathematics Department, University of Maine, Orono, ME 04473. To facilitate their consideration, solutions of Elementary Problems in this issue should be typed (with double spacing) and should be mailed before March 31, 1971. Contributors (in the United States) who desire acknowledgment of receipt of their solutions are asked to enclose self-addressed stamped postcards.

An asterisk (*) means neither the proposer nor the editors supplied a solution.

E 2265.* *Proposed by N. M. Dongre, Sydenham College, India*

Let a regular star polygon be constructed by dividing a circle into n equal parts and drawing the chords joining alternate points of division. Each of the n chords will carry four points of intersection. It is desired to assign the integers $1, 2, \cdots, 2n$ to the $2n$ points of intersection so as to have a magic star polygon (i.e. the sum of the four numbers on each chord is constant, see Problem E 2092 [1969, 557]). Prove that a necessary condition for the existence of a magic star polygon is that $n > 5$. Is this condition sufficient?

E 2266. *Proposed by Leon Gerber, St. John's University, Brooklyn, N.Y.*

If the lines joining the vertices of a simplex $\{A_i\}_0^n$ to its centroid M meet the circumsphere again in the points $\{B_i\}_0^n$, then

(a) $$A_iM\cdot MB_i = \sum A_iA_j^2/(n+1)^2, \qquad i,j = 0,\cdots,n, i\neq j.$$

(b) $$\sum A_iM/MB_i = n+1.$$

E 2267.* *Proposed by Emilia Mycielska, University of California, Berkeley*

Given a permutation $a_0, a_1, \cdots, a_n$ of the sequence $0, 1, \cdots, n$; a transposition of a_i with a_j is called *legal* if $a_i=0$ for $i>0$, and $a_{i-1}+1=a_j$. The permutation $a_0, a_1, \cdots, a_n$ is called *regular* if after a number of legal transpositions it becomes $1, 2, \cdots, n, 0$. For which numbers n is the permutation $1, n, n-1, \cdots, 3, 2, 0$ regular?

E 2268. *Proposed by Leonard Carlitz, Duke University*

An arithmetic function $f(n)$ is *completely multiplicative* if $f(ab)=f(a)f(b)$ for all positive integers a, b. Show that

$$\sum_{ab=n} f(a)f(b) = \delta(n)f(n) \qquad (n = 1, 2, \cdots),$$

where $\delta(n)$ is the number of divisors of n, if and only if $f(n)$ is completely multiplicative.

E 2269.* *Proposed by S. R. Conrad, Bayside, New York.*

A shuffled deck of ordinary playing cards is dealt out in the manner of the French gambling game of "treize" [cf. Rouse Ball, *Mathematical Recreations and Essays*, p. 325]. The dealer deals out the first 13 cards as he calls the numbers $1, 2, \cdots, 13$ [J, Q, K counting as 11, 12, 13, respectively]. He repeats the procedure three more times, without replacement, exhausting the deck. What is the probability of (a) no "match"? (b) exactly k matches? (c) at least k matches? [A match is defined as calling out the number n while dealing card n.]

E 2270. *Proposed by Peter Yff, American University of Beirut, Lebanon*

Let a, b, c be elements of a group G of odd order. If $a^2b^2=c^2$, prove that ab and c are in the same coset of the commutator subgroup K.

SOLUTIONS OF ELEMENTARY PROBLEMS

An Estimate of an Integral

E 2211 [1970, 79]. *Proposed by M. D. Landau, Lafayette College*

An exercise in Rudin, *Principles of Mathematical Analysis* states

$$\int_x^{x+1} \sin t^2 dt < 2/x \qquad \text{for } x > 0.$$

Can this bound be improved to some smaller multiple of $1/x$?

Solution by J. Gillis and M. Shimshoni, Weizmann Institute of Science, Israel. The bound can be replaced by $1/x$ and this is best possible. Indeed we show that

$$(1) \qquad \left| x\int_x^{x+1} \sin t^2 dt \right| < 1 \qquad \text{for } x > 0, \quad \text{and}$$

$$(2) \qquad \limsup_{x\to\infty} \left| x\int_x^{x+1} \sin t^2 dt \right| = 1.$$

Integration by parts yields

$$(*) \qquad \int_x^{x+1} \sin t^2 dt = \left[-\frac{\cos t^2}{2t} \right]_x^{x+1} - \frac{1}{2}\int_x^{x+1} \frac{\sin t^2}{t^2}\, dt,$$

and hence

$$\left| \int_x^{x+1} \sin t^2 dt \right| < \frac{1}{2}\left(\frac{1}{x} + \frac{1}{x+1}\right) + \frac{1}{2}\int_x^{x+1} t^{-2} dt = \frac{1}{x} \cdot$$

To prove (2) it will suffice to show that, given $\epsilon>0$, there exist arbitrarily large x such that $\cos x^2 > 1-\epsilon$, $\cos(x+1)^2 < -1+\epsilon$. It will then follow from (*) that

$$\left| \int_x^{x+1} \sin t^2 dt \right| > \frac{1-\epsilon}{2}\left[\frac{1}{x} + \frac{1}{x+1}\right] + O\left(\frac{1}{x^2}\right),$$

and so

$$\left| x\int_x^{x+1} \sin t^2 dt \right| > (1-\epsilon) + O\left(\frac{1}{x}\right).$$

But ϵ was arbitrary and x could be arbitrarily large, and hence

$$\limsup_{x\to\infty} \left| x\int_x^{x+1} \sin t^2 dt \right| \geqq 1.$$

This, combined with (1), yields (2). The existence of such x will follow if we show that, for any $\delta>0$, we can find positive integers m, n such that $|x_n^2 - 2m\pi| < \delta$, where x_n is defined by $(2x_n+1) = (2n+1)\pi$. For any n, let $y_n = (8\pi)^{-1}\{(2n+1)\pi - 1\}^2$ and we have to prove that the fractional part of y_n can be arbitrarily small. However, for any fixed positive integer h, we have $z_n = y_{n+h} - y_n = nh + \frac{1}{2}h[(h+1)\pi - 1]$ and the set $\{z_n\}$ is well known to be equidistributed, and the same is therefore true of the set $\{y_n\}$ by van der Corput's theorem [Acta. Math. 56(1931), 373].

Also solved by Meher Baba, D. M. Bloom, R. P. Boas, R. J. Dickson, W. O. Egerland, S. C. Ferry, Doug Hanto, Norman Locksley, O. P. Lossers (Netherlands), E. F. Schmeichel, Sid Spital, St. Olaf College Students, M. K. Vamanamurthy, Konrad Victor (Israel), L. E. Ward, Sr., and P. H. Young.

An Average Sequence

E 2212 [1970, 79]. *Proposed by G. E. Peterson, University of Missouri at St. Louis*

If $\{s_n\}$ and $\{t_n\}$ are sequences such that $\sum_{k=1}^{\infty} |t_k|$ converges and $\lim_{n\to\infty} (\sum_{k=1}^{n} s_k t_k)/s_n$ exists and is nonzero, then $\lim_{n\to\infty} s_n$ exists and is nonzero.

Solution by R. J. Dickson, Lockheed Palo Alto Research Laboratory. Set $r_n = (\sum_{k=1}^{n} s_k t_k)/s_n$; then $s_{k-1}r_{k-1}/s_k r_k = 1-(t_k/r_k)$. Let m be chosen large enough so that $r_n \neq 0$, $r_n \neq t_n$ for $n \geqq m$. Then $\prod_{k=m+1}^{n} (1-t_k/r_k) = s_m r_m / s_n r_n$. Letting $n \to \infty$, the conclusion follows if the product on the left converges. By a standard criterion, it suffices to have $\sum_{k=1}^{\infty} |t_k/r_k| < \infty$ and this follows from $\sum_{k=1}^{\infty} |t_k| < \infty$, since $r_n \to r \neq 0$.

Also solved by M. T. Bird, L. S. Bosanquet (England), Steven Ferry, M. L. J. Hautus (Netherlands), Ellen Hertz, D. E. Manes, E. F. Schmeichel, Simeon Reich (Israel), Sid Spital, Mark Yu, and the proposer.

Quadrilaterals with the Nagel Property

E 2213 [1970, 79]. *Proposed by H. Demir, Middle East Technical University, Turkey*

Let us say that a (planar) polygon has the *Nagel property* if the lines through the vertices of the polygon and bisecting the perimeter of the polygon are concurrent. It is known that all triangles have the Nagel property and that not all quadrilaterals have the property. Determine the simple nondegenerate quadrilaterals that have the Nagel property.

Solution by the editor based on the proposer's solution. Let $ABCD$ be a quadrilateral having the Nagel property. Let each of the lines AA', BB', CC', DD' bisect the perimeter and pass through the Nagel point N. Let $AB=a$, $BC=b$, $CD=c$, $DA=d$. We may suppose $a+b \geqq c+d$ and $b+c \geqq a+d$. It then follows that C' lies on segment AB, B' on CD, and A' and D' on BC (Fig. 1). Set $BD'=u$,

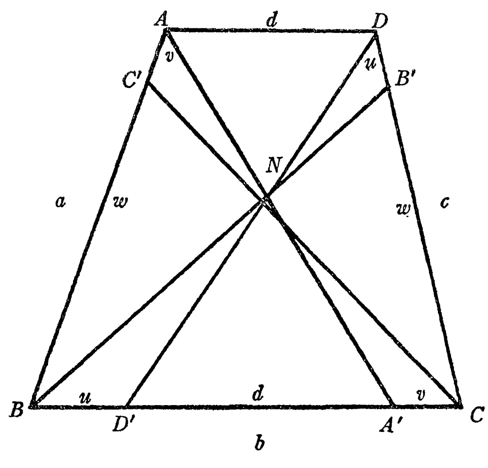


Fig 1

$A'C=v$, $CB'=w$, and $a+b+c+d=2s$. Since $BC+CB'=s=D'C+CD$, then $DB'=u$. Similarly $AC'=v$, $BC'=w$ and $D'A'=d$. We distinguish three cases.

CASE 1. $u\neq 0$ and $v\neq 0$. Thinking of AA' and DD' meeting at N, we see that a necessary and sufficient condition for CC' to pass through N is

$$\frac{v}{w}\cdot\frac{b}{-v}\cdot\frac{A'N}{NA}=-1$$

by Menelaus' theorem applied to triangle ABA' cut by line CNC'. Similarly, BB' passes through N if and only if

$$\frac{-u}{b}\cdot\frac{w}{u}\cdot\frac{DN}{ND'}=-1,$$

using triangle $DD'C$ cut by BNB'. Then

$$\frac{DN}{ND'}=\frac{b}{w}=\frac{AN}{NA'}\cdot$$

Hence AD is parallel to $D'A'$, and since these segments are also equal, it follows that $ADA'D'$ is a parallelogram. The diagonals AA' and DD' then bisect each other, so $AN=NA'$ and $b=w$. But $b+w=s$, so $2w=a-b+c+d=2b$, from which we obtain

$$b=\frac{a+c+d}{3}\cdot$$

Hence $ABCD$ is a trapezoid such that the longer of its two parallel bases is the arithmetic mean of its other three sides. Reversing the argument of this paragraph shows that every such trapezoid has the Nagel property. For example, the trapezoid with vertices (0, 0), (19, 0), (6, 24), (0, 24) has $N=(7, 12)$.

CASE 2. $u=v=0$. Then $a+b=c+d=s$ and $a+d=b+c=s$, so $2a+b+d=2c+b+d$. Hence $a=c$ and $b=d$, so $ABCD$ is a parallelogram. Clearly every parallelogram has the Nagel property.

CASE 3. $u=0$ and $v\neq 0$ (or vice versa). Let AC and BD meet at M (Fig. 2). Now $a+d=b+c=s$, so $d+v+w=C'B+BC=CD+DA+AC'=c+d+v$, whence $c=w$. Applying Menelaus' theorem to triangle ABA' cut by CNC' and to triangle $AA'C$ cut by BNM, we obtain

$$\frac{b}{-v}\cdot\frac{A'N}{NA}\cdot\frac{v}{c}=-1,\qquad \text{so } \frac{A'N}{NA}=\frac{c}{b},$$

$$\frac{AN}{NA'}\cdot\frac{-d}{b}\cdot\frac{CM}{MA}=-1,\qquad \text{so } \frac{CM}{MA}=\frac{b}{d}\cdot\frac{A'N}{NA}=\frac{b}{d}\cdot\frac{c}{b}=\frac{c}{d}\cdot$$

Since M divides side CA of triangle DAC in the ratio c/d of the adjacent sides, then DM bisects angle D. Hence $\measuredangle ADM=\measuredangle MDC=\alpha$. Applying the law of

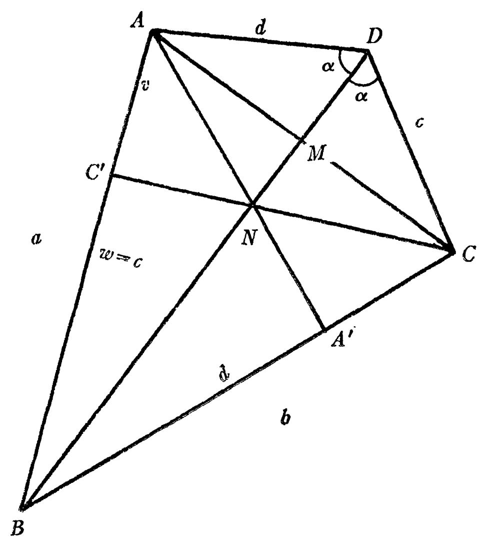


FIG. 2

cosines to triangles ABD and CBD and letting $BD=f$, we have

$$a^2 = d^2 + f^2 - 2df \cos \alpha,$$
$$b^2 = c^2 + f^2 - 2cf \cos \alpha,$$

so

$$a^2 - b^2 = d^2 - c^2 - 2f(d - c) \cos \alpha.$$

Assuming $d \geqq c$, let $d = c+e$, so $b = c+v+e$ and $a = c+v$. Now $a^2-(a+e)^2 = (c+e)^2 - c^2 - 2ef \cos \alpha$, which simplifies to

$$ef \cos \alpha = ce + ae + e^2.$$

If $e \neq 0$, then $\cos \alpha = (c+a+e)/f = (c+b)/f$, an impossibility since $c+b > f$. Hence $e=0$, so $c=d$ and $a=b$. The figure is therefore a kite. By symmetry, every kite has the Nagel property.

Three other correspondents mentioned the kite and the parallelogram.

Interior Diagonals of a Polygon

E 2214 [1970, 79]. *Proposed by M. S. Klamkin, Ford Scientific Laboratory and B. Ross Taylor, York High School.*

It is intuitive that every simple n-gon ($n > 3$) possesses at least one interior diagonal. For a simple n-gon what is the least number of diagonals which, except for their endpoints, lie wholly in its interior?

Solution by Anders Bager, Hjørring, Denmark. The two tangents from a point P outside a circle Γ touch Γ in points A and B. Connect A and B with a broken line consisting of $n-2$ chords succeeding each other along the smaller arc from A to B. Join P to A and B to obtain a simple n-gon with exactly $n-3$ inner diagonals (all issuing from P).

The number $n-3$ is minimal. This is trivially so if $n=3$. Suppose it true for some n and consider an arbitrary simple $(n+1)$-gon. From this cut off a triangle such that two sides are sides of the $(n+1)$-gon, and the third side an inner diagonal. This is always possible and leaves a simple n-gon which, by assumption, has at least $n-3$ inner diagonals. Hence the $(n+1)$-gon has at least $(n-3)+1=(n+1)-3$ inner diagonals. Thus the assertion of the problem is true by induction.

Also solved by M. Deakin (New Guinea), R. B. Eggleton (Australia), Michael Goldberg, M. G. Greening (Australia), Norman Miller, E. F. Schmeichel, Roger Weitzenkamp, William Wernick, W. G. Wild, Joseph Zaks, and the proposers.

Eggleton establishes the result that a simple n-gon has precisely $n-3$ inner diagonals if and only if no two of its diagonals intersect.

Packing Circles into Quadrilaterals

E 2215 [1970, 192]. *Proposed by Michael Goldberg, Washington, D.C.*

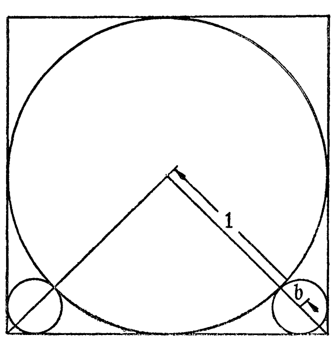
Find the shapes of the quadrilaterals of least area K which enclose three given circles of radii $1, b, b$.

Solution by Thomas Hughes, Arlington, Texas. The formulas for K as a function of b are as follows (compare the corresponding figures):

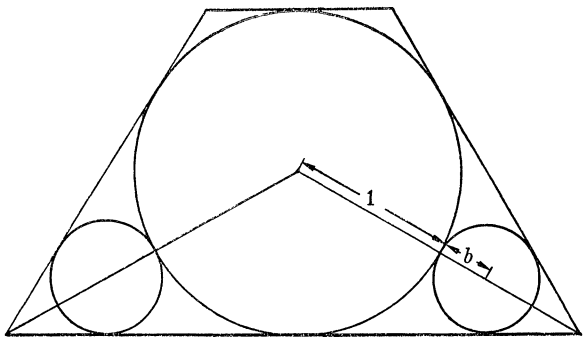
(1) $$0 \leqq b \leqq (\sqrt{2}-1)/(\sqrt{2}+1), \qquad K = 4.$$

(2) $$(\sqrt{2}-1)/(\sqrt{2}+1) \leqq b \leqq 0.415, \qquad K = (1+b)^2/(\sqrt{b}(1-b)).$$

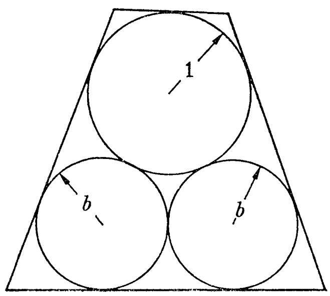
(3) $0.415 \leqq b \leqq 0.926,$
$$K = (1+b+\sqrt{1+2b})((1+b+\sqrt{1+2b}) \cot A + 2\cot(90^\circ - \tfrac{1}{2}A))$$



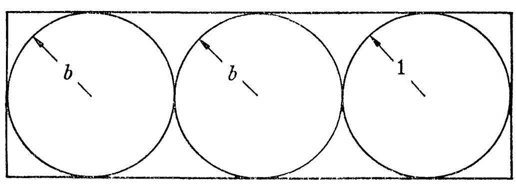
(1)



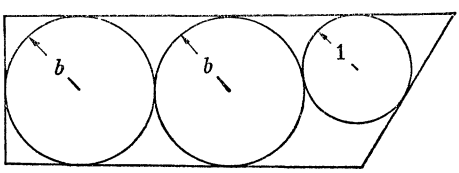
(2)



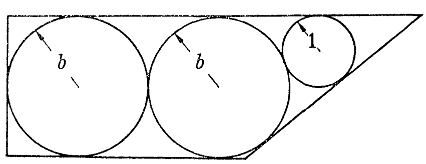
(3)



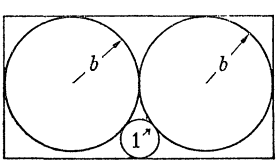
(4)



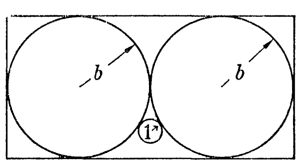
(5)



(6)



(7)



(8)

where $A = B+C$, $B = \text{Arcsin}((1-b)/(1+b))$, $C = \text{Arccos}(b/(1+b))$.

(4) $$0.926 \leqq b \leqq 1, \qquad K = (d+b+1)\tan\theta + b\cot\theta,$$

where $d = (5b^2+8b\sqrt{b}+2b+1)^{1/2}$, $\theta = 45° + \frac{1}{2}\text{Arcsin}((1-b)/d)$.

(5) $$1 \leqq b \leqq 2\sqrt{2}-1, \qquad K = 2b(3b+2\sqrt{b}+(b(2-b))^{1/2}).$$

(6) $2\sqrt{2}-1 \leqq b \leqq 3.783,$
$$K = b(6b+4\sqrt{b}+2\sqrt{b}(2-b)/(b-1)+b(b-1)/2\sqrt{b}).$$

(7) $$3.783 \leqq b \leqq 4, \qquad K = 4b(b+2\sqrt{b}).$$

(8) $$4 \leqq b, \qquad K = 8b^2.$$

Also solved by the proposer.

A Comparison of Integrals

E 2216 [1970, 192]. *Proposed by M. S. Klamkin, Ford Scientific Laboratory*

Which of the two integrals

$$\int_0^1 x^x dx, \qquad \int_0^1 \int_0^1 (xy)^{xy} dx\, dy$$

is larger?

Solution by R. A. Groeneveld, Mount Holyoke College. Making the substitution $u = xy$, the second integral may be written

$$\int_0^1 \int_u^1 \frac{u^u}{x} dx du = -\int_0^1 u^u(\log u) du.$$

Since

$$\int_0^1 u^u(1+\log u) du = u^u \Big|_0^1 = 0,$$

the two stated integrals are equal.

Also solved by D. S. Adorno, Christian Andersen (Denmark), K. F. Anderson, J. Beeman & B. Dimsdale, J. W. Burns, Joe Chance, J. A. Coatimundi, Steven Ferry, N. J. Fine, B. A. Fusaro, J. Gillis (Israel), M. F. Gillis, David Gootkind, C. B. Grosch, Judith R. Gumerman, Seymour Haber, Ellen Hertz, R. A. Horn, D. G. Kabe, M. M. Klein, E. F. Knapp, H. S. M. Kruijer (Netherlands), J. R. Kuttler, Harry Lass, J. J. Leeson & D. L. Sherry, H. S. Lieberman, J. G. Mauldon, Simeon Reich (Israel), Valerie Rindone, E. F. Schmeichel, Michael Shimshoni (Israel), J. S. Shipman, G. A. Stoops, St. Olaf College Students, E. W. Trost (Switzerland), M. K. Vamanamurthy, M. M. Wells, J. E. Wilkins, Jr., and the proposer.

J. Gillis proves the following generalization: Define

$$I_r = \int_0^1 \int_0^1 \cdots \int_0^1 (x_1x_2\cdots x_r)^{x_1x_2\cdots x_r} dx_1 dx_2 \cdots dx_r,$$

$r = 1, 2, \cdots$. Then $I_1 = I_2 < I_3 < I_4 < \cdots$, and $\lim_{r\to\infty} I_r = 1$. Klein reports the computer value of I_1 is 0.78343051.

ADVANCED PROBLEMS

All solutions of Advanced Problems should be sent to J. Barlaz, Rutgers—The State University, New Brunswick, NJ 08903. Solutions of Advanced Problems in this issue should be typed (with double spacing) on separate, signed sheets and should be mailed before March 31, 1971. Contributors (in the United States) who desire acknowledgement of receipt of their solutions are asked to enclose self-addressed, stamped postcards.

An asterisk (*) means neither the proposer nor the editors supplied a solution.

5765.* *Proposed by Simeon Reich, Israel Institute of Technology, Haifa*

On p. 325 of Dugundji, *Topology*, the following result is stated: A is a deformation retract of B over X if and only if A is a retract of B and B is deformable into A over X.

Is this true?

5766. *Proposed by Morris Newman, National Bureau of Standards*

Let α, β be complex numbers linearly independent over the reals. Let $g(z)$ be an entire function, and suppose that there are constants a, b such that for all z, $g(z+\alpha)=ag(z)$, $g(z+\beta)=bg(z)$. Prove that constants A, B exist such that $g(z)=A\exp(Bz)$.

5767. *Proposed by Howard Jacobowitz, New York University*

Let $g(t)$ be a real valued function on the open unit interval. The Cauchy criterion can be expressed in the following form:

$$\lim_{t\to 0}\sup_{0<t_1,t_2<t}\{g(t_1)-g(t_2)\}=0$$

implies that $g(t)$ has a finite limit as $t\to 0$.

Prove or disprove the following modification:

$$\lim_{t\to 0}\sup_{t<t_1,t_2<2t}\{g(t_1)-g(t_2)\}=0$$

implies that $g(t)$ has a finite limit as $t\to 0$.

5768.* *Proposed by Peter Flor, University of Vienna, Austria*

Minkowski's singular monotone function $M(x)$ is defined as follows: $M(0)=0$, $M(1)=1$; if $a=p/q$ and $b=p'/q'$ are rational numbers in $[0, 1]$ such that $pq'-p'q=1$, and if $c=(p+p')/(q+q')$, then $M(c)=\frac{1}{2}[M(a)+M(b)]$. This defines $M(x)$ for every rational $x\in[0, 1]$; the function can then be extended by continuity to all of $[0, 1]$. Obviously 0, $\frac{1}{2}$, and 1 are fixed points of $M(x)$. Prove that there are exactly two further fixed points, $d_1=0.4203723\cdots$ and $d_2=1-d_1$. Decide whether they are rational.

5769. *Proposed by L. W. Shapiro, Howard University*

Show that a finite simple group has no irreducible representation over the complex numbers of degree two.

5770.* *Proposed by J. P. Jones, University of Calgary*

Does there exist a function $\psi=\psi(x, y)$ continuous on some domain in R^2 with the following property? For each real valued $\theta(x)$ continuous on a connected domain, there exists a y_0 such that $\psi(x, y_0)$ and $\theta(x)$ have the same domain and agree there.

SOLUTIONS OF ADVANCED PROBLEMS

Product of Algebraic Numbers

5674 [1969, 565; 1970, 661]. *Proposed by L. Carlitz, Duke University*

It is proved in problem 5542 [1968, 1021] that the following statement is incorrect: If a, b are algebraic over F of degree m and n respectively and if m and n are relatively prime, then ab is algebraic over F of degree mn.

Show that the statement is correct when $F=GF(q)$, $q=p^t$, p prime, $t\geqq 1$.

II. *Solution by R. A. Moore, Southern Illinois University.* Let F be any field with $F(a)$ and $F(b)$ normal separable extensions of F of degrees m and n respectively. If $(m, n)=1$, then Galois Theory yields the following facts:

(i) $\deg(F(a, b)/F)=m\cdot n$.

(ii) The Galois groups $G(F(a, b)/F(a\cdot b))$ and $G(F(b)/F(b)\cap F(a\cdot b))$ are isomorphic.

It follows that $\deg(F(a, b)/F(a\cdot b))$ divides $\deg(F(b)/F)=n$.

In like manner, $\deg(F(a, b)/F(a\cdot b))$ divides m. Since $(m, n)=1$, $F(a, b)=F(a\cdot b)$. This includes the special case $F=GF(q)$.

Editorial Note. The author of Solution I [1970, 661] reports that his solution is incorrect. Irving Guest has provided a counterexample to the claim that ab is a root of unity of order l.c.m. (α, β).

A Set Hierarchy

5697 [1969, 1074]. *Proposed by Frederick Hammer, Paine College, Augusta, Georgia*

A set is *transitive* if $x\in y\in S$ implies $x\in S$, disjoint if $x, y\in S$ implies $x\cap y=\varnothing$ or $x=y$. The *Axiom of Regularity* states that $x\neq\varnothing$ implies the existence of $y\in x$ with $y\cap x=\varnothing$. Show that in each finite cardinality there is exactly one transitive and disjoint set, and further, that there is only one transitive and disjoint infinite set, assuming the Axiom of Regularity.

Solution by W. G. McArthur, Shippensburg (Pa.) State College. Let $N_1=\varnothing$ and for k a natural number greater than or equal to 2, let $N_k=\{N_{k-1}\}$. Let $A_0=\varnothing$, $A_1=\{N_1\}$, and in general, $A_k=\{N_1, N_2, \cdots, N_k\}$. Finally, let $A_\infty=\bigcup_{k=1}^{\infty}A_k$. Evidently each A_k and A_∞ is transitive and disjoint. Furthermore, card $A_k=k$ and card $A_\infty=\aleph_0$.

We shall first prove, by induction, that if S is a transitive disjoint set and

card $S \geqq k$, then $N_k \in S (k \geqq 1)$. For $k=1$, suppose $N_1 = \varnothing$ is not an element of S. By the Axiom of Regularity, there is an $x \in S$ such that x and S are disjoint. Since $x \neq \varnothing$, there is $y \in x$. By transitivity of S, y is also an element of S. Thus, y is in $S \cap x$, a contradiction, so that $N_1 \in S$. Now assume that the condition is true for a fixed k and suppose that card $S \geqq k+1$. By the inductive hypothesis, $A_k \subset S$. Let $T = S - A_k$ and note that $T \neq \varnothing$. Thus, by the Axiom of Regularity, there is $x \in T$ such that $x \cap T = \varnothing$. Since $x \neq \varnothing$, there is $y \in x$. Then $y \in S - T$ and hence $y \in A_k$. By disjointness of S, $y = N_k$, and we must have $x = \{N_k\} = N_{k+1}$.

Now, let S be a finite transitive disjoint set with card $S = k$; by the preceding we have $S = A_k$. On the other hand, if S is an infinite transitive disjoint set, then $A_\infty \subset S$. If $A_\infty \neq S$, then by the Axiom of Regularity, there is an element x in $S - A_\infty$ such that x and $S - A_\infty$ are disjoint. Then $x \neq \varnothing$ and there is a y in x; by transitivity, y is in A_∞. Hence there is a k such that $y = N_k$. But $\{N_k\} = N_{k+1} \in A_\infty \subset S$. This contradicts the disjointness of S. Thus $S = A_\infty$ and the proof is complete.

Also solved by M. L. Laplaza (Puerto Rico), J. M. Plotkin, and the proposer.

Inner R-Derivations

5707 [1970, 84]. *Proposed by W. A. Vasconcelos, Rutgers—The State University*

Let R be an integral domain and G a finite group. Assume the characteristic of R does not divide $|G|$. Prove that each R-derivation of $R[G]$ is inner.

Solution by Barbara Osofsky, Rutgers—The State University. Let δ be an R-derivation of $R[G]$ and let $\{c_i | 1 \leqq i \leqq n\}$ be the distinct conjugate classes of G, $h_i \in c_i$, $d_i = \sum_{g \in c_i} g \in R[G]$. Let $x = \sum_{h \in G} h\delta(h^{-1}) = \sum_{g \in G} \alpha_g g$, $\alpha_g \in R$. Let $y = x - \sum_1^n \alpha_{h_i} d_i$. Then x and y induce the same inner derivation, and the coefficient β_{h_i} in $y = \sum_{g \in G} \beta_g g$, $\beta_g \in R$, must be 0.

Now, for all $g \in G$,

$$g^{-1}xg = \sum_{h \in G} g^{-1}h\delta(h^{-1})g = \sum_{h \in G} g^{-1}h[\delta(h^{-1}g) - h^{-1}\delta(g)]$$

$$= \sum_{h \in G} g^{-1}h\delta(h^{-1}g) - \sum_{h \in G} g^{-1}\delta(g) = x - |G| g^{-1}\delta(g).$$

So $|G|\delta(g) = gx - xg = gy - yg$ for all $g \in G$.

Let $k \in G$. $k = gh_i g^{-1}$ for some g and i. Then the coefficient of gh_i in gy is zero, and the coefficient of gh_i in $|G|\delta(g)$ is divisible by $|G|$, so that the coefficient of $gh_i = kg$ in yg, namely β_k, is divisible by $|G|$. Hence δ is the inner derivation induced by $(1/|G|)y \in R[G]$.

Also solved by G. A. Elliott, E. R. Gentile (Argentina), M. G. Greening (Australia), Miguel Torres (Spain), and the proposer.

Separable Extension Fields

5708 [1970, 84]. *Proposed by C. W. Avery, San Jose State College*

Let K be a finite extension of the field k, complete in a non-Archimedean valuation. Let $\bar{K}$ and $\bar{k}$ denote the residue class fields. Problem 16, p. 129 of P. J. McCarthy, *Algebraic Extensions of Fields* asserts that K is separable over k if $\bar{K}$ is separable over $\bar{k}$. Disprove.

Solution by Robert Gilmer, Florida State University. Let F be a field of characteristic $p\neq 0$, let X be an indeterminate over F, let K be the quotient field of the formal power series ring $F(X)[[Y]]$, and let k be the quotient field of $F(X)[[Y^p]]$. If v is a valuation on K associated with the valuation ring $F(X)[[Y]]$, and if w is similarly defined on k for $F(X)[[Y^p]]$, then K and k are complete under v and w, respectively. However, K/k is purely inseparable of degree p, while $\bar{K}=\bar{k}=F(X)$.

Also solved by Louis Dernier, Harley Flanders (Israel), Harald Niederreiter, T. F. Zelman, and the proposer.

A Linear Combination of Normally Distributed Variables

5709 [1970, 85]. *Proposed by W. A. J. Luxemburg, California Institute of Technology*

For all $x>0$, determine

$$\lim_{n\to\infty} \frac{1}{(\sqrt{\pi})^n}\int_{D_n(x)}\cdots\int \exp[-(x_1^2+\cdots+x_n^2)]dx_1\cdots dx_n,$$

where

$$D_n(x) = \left\{(x_1,\cdots,x_n): \left|\frac{x_1}{1}+\frac{x_2}{\sqrt{2}}+\cdots+\frac{x_n}{\sqrt{n}}\right| \leqq x\right\}.$$

Solution by J. C. Tanner, Crowthorne, Berkshire, Great Britain. For a fixed n, the problem may be reformulated as follows: If $x_1, x_2, \cdots, x_n$ are independently and normally distributed with zero mean and variance $\frac{1}{2}$, what is the probability that

$$\left|\frac{x_1}{1}+\frac{x_2}{\sqrt{2}}+\cdots+\frac{x_n}{\sqrt{n}}\right| \leqq x?$$

Denoting the left hand side by $|y|$, y is normally distributed with zero mean and variance

$$\sigma_n^2 = \frac{1}{2}\left(1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n}\right),$$

and therefore the probability that $|y|\leqq x$ is $2F(x/\sigma_n)-1$, where

$$F(X) = \int_{-\infty}^{X} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

is the normal integral. As n tends to infinity, σ_n tends to infinity, and the probability tends to $2F(0)-1=0$.

Also solved by D. R. Breach (New Zealand), L. E. Clarke (England), Michael Deakin (New Guinea), E. J. Dudewicz, Peter Ennis, Joe Flowers, David Gootkind, J. R. Gumerman, D. A. Hejhal, N. L. Johnson, Morton Kupperman, O. P. Lossers (Netherlands), M. F. Neuts, G. S. Rogers, E. F. Schmeichel, and the proposer.

A Special Topology for the Integers

5712 [1970, 85]. *Proposed by Dan Marcus, York University, Toronto*

Is it possible to topologize the integers in such a way that the connected sets are the sets of consecutive integers? Generalize to the lattice points of n-space.

Solution by Cleveland State University Problem Solving Group, Frank Wyse, Advisor. Let L denote the set of lattice points in n-space. For x, y in L, we write $x\mathrm{R}y$ if and only if they differ by 1 in one coordinate and agree in all of the others. For x, y elements of a subset $S\subseteq L$, we write $x\mathrm{E}_s y$ if and only if there is a sequence $x=x_1, x_2, \cdots, x_n=y$ (where n may be any positive integer) such that $x_i\in S$ for each index i and $x_i\mathrm{R}x_{i+1}$ for $i<n$. We say $S\subseteq L$ is a set of consecutive points of L if $x\mathrm{E}_s y$ whenever x and y are in S.

Now for any $x\in L$, let $U_x=\{y\in L: y=x \text{ or } y\mathrm{R}x\}$ if the sum of the coordinates of x is even; and let $U_x=\{x\}$ otherwise. Then $x\neq y$ and z in $U_x\cap U_y$ imply $U_z=\{z\}$, so that $\{U_x: x\in L\}$ is a base for a topology on L such that each $x\in L$ has a smallest neighborhood, namely U_x. Let L have this topology.

Now suppose C is a nonempty connected subset of L. Choose $x_0\in C$, let $A_1=\{y\in C: y\mathrm{E}_c x_0\}$, and let $A_2=C-A_1$. Then $x\in A_i$ implies $U_x\cap C\subseteq A_i$, so that A_1 and A_2 are both open in C. Hence $A_2=\varnothing$, and $C=A_1$ is a set of consecutive points of L.

Conversely, suppose C is a set of consecutive points of L. Let A and B be disjoint nonempty sets whose union is C. Choose $x\in A$, $y\in B$ and a sequence $x=x_1, x_2, \cdots, x_n=y$ such that $x_i\in C$ for all i and $x_i\mathrm{R}x_{i+1}$ for $i<n$. Then, for some i we have $x_i\in A$ and $x_{i+1}\in B$, so that either $x_i\in U_{x_{i+1}}$ in which case B is not open in C, or $x_{i+1}\in U_{x_i}$ in which case A is not open in C. Hence, C is connected.

Also solved by M. E. Ballotti, J. L. Bryant & R. W. Gilmer, F. A. Delahan, Ellen Hertzmark, H. C. Kranzer, Michael McCoy, P. J. Saavedra, E. F. Schmeichel, Konrad Victor (Israel), and the proposer.

INDEX TO VOLUME 77, 1970

THE AMERICAN MATHEMATICAL MONTHLY

AUTHOR INDEX

SUBJECT INDEX

CLASSIFICATION

1. GENERAL

2. MATHEMATICAL EDUCATION

3. FOUNDATIONS AND SET THEORY

4. COMBINATORIAL ANALYSIS

5. NUMBER THEORY

6. LINEAR ALGEBRA

7. ALGEBRA

8. REAL ANALYSIS

9. COMPLEX ANALYSIS

10. DIFFERENTIAL EQUATIONS

11. ABSTRACT ANALYSIS

12. GEOMETRY

13. TOPOLOGY

14. PROBABILITY AND STATISTICS

15. COMPUTER SCIENCE

16. APPLIED MATHEMATICS

PROBLEMS AND SOLUTIONS

PROBLEMS PROPOSED

PROBLEMS SOLVED

SOLUTIONS

Numbers in **boldface** type refer to problems, those in lightface, to pages.

REVIEWS

Abbreviations: (TR)—Telegraphic Review; (NP)—Notable Paper.
Names of authors are in ordinary type, those of reviewers in capitals.

FILM REVIEWS

NEWS AND NOTICES

PERSONAL ITEMS

GENERAL INFORMATION

NECROLOGY

REPORTS AND ANNOUNCEMENTS OF THE ASSOCIATION AND ITS SECTIONS

MEETINGS AND ANNOUNCEMENTS OF THE ASSOCIATION

MEETINGS OF ITS SECTIONS